

## Missing the big picture

**Our understanding of the likely ecological impact of genetically modified crops is incomplete. But these holes in our knowledge are symptomatic of a wider failure adequately to address the science of sustainable agriculture.**

**G**M, or not GM? That is the question ... or, at least, that is an important question with which the British government must wrestle over the next few months, and the answer to which will have international repercussions.

More than three years ago, with agribiotech companies pushing to market crops engineered to resist the effects of broad-spectrum herbicides, and UK public opposition to genetically modified (GM) crops reaching feverish levels, Prime Minister Tony Blair's administration delayed a difficult decision. There would, the government said, be no commercial planting of GM crops until the effects on farmland biodiversity of the herbicide applications associated with the crops' cultivation had been evaluated through extensive ecological trials. By the time these were completed, the politicians must have hoped, public opposition would have ebbed away.

Some hope. The results of these farm-scale trials are now being written up, and although the frenzy of 1999 has subsided, the British public shows little sign of warming to GM agriculture. Like it or not, Britain's decision on whether to allow commercial planting of GM herbicide-tolerant crops will be seen as a verdict on the wider future of UK transgenic agriculture. Even if, as seems likely, the farm-scale trials give herbicide-tolerant crops a relatively clean bill of health with regard to biodiversity, they can say little about the benefits and risks of GM agriculture as a whole, as the government's own Agriculture and Environment Biotechnology Commission reported in September 2001.

So, belatedly, Blair's government has launched a wider exercise to consider the merits of transgenic farming. A public debate, including a series of meetings and perhaps a specially commissioned film, is planned. Two expert panels have also been set up, one to assess the economic arguments, the other to review the scientific evidence.

The latter panel, which is to report in the early summer, must review the literature and perhaps unpublished studies, and address questions posed by the public over the Internet. Its report will be an influential document, and not just in Britain. GM agriculture has been accepted in North America, but many countries have yet to embrace the technology. And the developing world, already a battleground for pro- and anti-GM lobbyists (see page 681), is bound to look for guidance to expert panels in leading scientific nations.

### Trial run

Anyone surveying the literature with an unbiased eye should conclude that, after years of investigation, there is no convincing evidence that GM crops pose risks to human health, or that they will lead to an ecological meltdown. The farm-scale trials may also provide reassurance that herbicide-tolerant GM crops can be grown without adverse effects on farmland biodiversity. And a study published last month, backed by the agribiotech giant Monsanto, claimed that such crops can even boost invertebrate populations, if combined with specific regimes of herbicide application (A. M. Dewar *et al. Proc. R. Soc. Lond. B* 270, 335; 2003).

Yet Britain's GM science panel should also acknowledge that the research is lacking in several respects. Take the issue of gene flow from GM crops to wild relatives. Many studies, trumpeted by anti-GM activists at every opportunity, have shown that transgenes can

spread beyond the crops from which they were introduced. But despite industry's past tendencies to play down the possible extent of gene flow, we have long known that crops will hybridize with related weeds. The real issue is whether the flow of transgenes has any undesirable ecological or agronomic consequences.

Answering this question will involve creating hybrids between transgenic crops and wild relatives, and monitoring their effects on farmland ecology and crop yields when released in field experiments. Such trials are now under way in North America; one preliminary example, which delivered a reassuring message, was the highlight of a conference on transgene flow in Amsterdam last month (see *Nature* 421, 462; 2003). But European regulators have not bitten this bullet. A decade ago, they ignored proposals by some far-sighted ecologists to study the consequences of transgene flow. Now they are running scared of public opinion, which has been primed by anti-GM activists to see such trials themselves as inherently risky.

The necessary experiments can be done using male-sterile plants, which shouldn't breed and pose any lasting hazard. But in Europe, GM crops have been demonized so effectively that it is almost impossible to carry out the research to determine whether fears about invasive 'superweeds' have any foundation. Breaching this impasse won't be easy. Some proponents of transgenic agriculture claim that the risks are small, and argue that we should push ahead with commercial plantings. This just isn't good enough. Dismissing legitimate public concerns will only harden opposition to transgenic farming.

### Broader view

Ultimately, the answer has to involve placing the arguments about GM crops in a wider context. Meeting the nutritional needs of the world's growing population while protecting the planet's biodiversity is a huge challenge. To meet it, we can ill afford to cast aside entire technologies without testing whether they can be effectively and safely deployed. This applies to transgenics, but also to enhancements to conventional breeding allowed by our growing genomic and molecular-genetic knowledge (see *Nature* 421, 568–570; 2003).

Shamefully, when it comes to creating new varieties that might help to feed the developing world's growing population, rich countries are now cutting spending on both approaches to crop improvement. And amid all the fuss about GM crops, there's been little acknowledgement that similar questions about biodiversity and gene flow must be asked about conventionally bred varieties. Take a variety of rice that can tolerate saline conditions. Such a crop, created by transgenic or conventional means, would allow the cultivation of soils that are now seen as agricultural wastelands. But might it also spawn superweeds that would choke estuarine habitats? Such questions need answers. At present, however, there seems to be little desire to find them.

Britain's farm-scale trials of herbicide-tolerant GM crops represent an unprecedented effort to study the ecological impact of a change in agricultural practice. They could serve as a blueprint for experiments to study a whole range of farming practices, putting sustainable agriculture on a sound scientific footing. But scientists, regulators and politicians must seize the initiative and widen the debate about the future of farming beyond an obsession with transgenics. ■





**Variation game**  
Study generates mixed response for Icelandic genes  
**p678**



**One for the road**  
The brakes are off for London's traffic congestion toll  
**p679**



**Top crop or flop?**  
Experts disagree over success of India's first GM harvest  
**p681**



**In a stew**  
Have transgenic pigs escaped into the US food supply?  
**p682**

# Shuttle inquiry to piece disaster together from the ground up

**Declan Butler**

The investigation into the Columbia disaster is focusing on a huge reverse-engineering analysis. NASA staff are working to pinpoint where and when each of the 12,000 burnt and twisted pieces of debris landed. The trajectories of these pieces, which are scattered over some 70,000 square kilometres of Texas and Louisiana, will then be plotted to create a virtual video of the craft breaking up.

Known as ballistic trajectory analysis, the technique is among the panoply of forensics being used in the crash investigation by NASA and an independent inquiry board, chaired by retired US Navy admiral Harold Gehman. The investigators will map the coordinates of each piece of wreckage using the Global Positioning System, and estimate its trajectory with the help of data on wind speed and the weight and shape of each part. This should shed light on the sequence in which key parts of Columbia broke off — those that fell first are most likely to be related to the cause of the crash.

Such techniques are used to analyse airline crashes. But the high altitude and speed — 60 kilometres and more than 20,000 kilometres per hour — at which Columbia broke up complicate matters. The debris is strewn over a much wider area than the few square



**Parts and labour:** Harold Gehman (fifth from left) will study debris to reconstruct Columbia's demise.

miles typical in an aircraft crash, says William Waldo, a flight-crash analyst at Embry-Riddle Aeronautical University in Prescott, Arizona. Little accurate data on wind speed are available for altitudes above 15 kilometres, he adds.

Useful information may also be extracted from individual pieces of debris by forensic scientists at Barksdale Air Force Base in Louisiana, where all of the debris is being assembled. Electron-microscope images of the grain structure of metal debris will reveal

## Biologists wary that cash up front could mean cuts later

**Erika Check, Washington**

A new way of allocating grants at the US National Institutes of Health (NIH) could leave the agency vulnerable to budget cuts, biology researchers have warned.

President George Bush's 2004 budget request, made last week, asks the NIH to award 322 new research grants using a "fully funded" mechanism, under which all the money for the grant would be paid at the start. The NIH usually splits funding into a series of yearly payments made over the lifetime of a grant.

Fully funded grants are aimed at companies in the private sector, which may be deterred by the possibility that

their grants could be cut in the years after they are awarded, says Donald Poppke, the NIH's acting associate director for budget. The request also asks the NIH to study whether full funding works, and whether more NIH grants should be paid in this way.

Observers say that researchers might appreciate the certainty that comes with a fully funded grant, but worry that it could make it easier for the government to cut the NIH's budget in subsequent years. "The NIH is potentially vulnerable to cuts if there is no commitment to fund grants beyond their first year," says

Bob Rich, executive associate dean at Emory University School of Medicine in Atlanta, Georgia.

Steven Teitelbaum, president of the Federation of American Societies for Experimental Biology, points out that this could be particularly difficult in tight fiscal times, such as those being experienced now, when the NIH will be awarding fewer grants than in the immediate past (see *Nature* 421, 565; 2003). But for the moment, the federation and other advocacy groups have not taken a position on the issue. "It's a new kid on the block," says Teitelbaum "We're going to have to think about it."

▶ tell-tale signs of the heating that each part experienced, says Arvid Pasto, director of the High Temperature Materials Laboratory at Oak Ridge National Laboratory in Tennessee. Knowing the shape and conductivity of the shuttle's components will also allow the spreading of heat throughout the craft to be modelled. And if pieces are missing, they probably either blew apart or were vaporized, again giving clues as to where fires broke out.

NASA is likely to send debris to Oak Ridge for such tests, says Pasto. Parts that experienced the most heating were probably situated close to where the malfunction originated. But the ultimate goal, he says, is to simulate the heat to which all of the individual pieces were exposed and then reconstruct the entire break-up with the aid of a supercomputer.

NASA staff will check their break-up models with radar and satellite images of the disintegrating shuttle and the wealth of photos and videos taken by the public and astronomers. They will also make use of data from the shuttle's autopilot, which took corrective action to counter a drag on the left wing minutes before the craft broke up. Models of the shuttle's aerodynamics could shed light on what caused the autopilot to take the action that it did.

Little other concrete information about the moments before the accident has emerged so far. Just before contact with the craft was lost, mission control detected temperatures rising by around 2 °C per minute in the well that contains the wheel on the left wing. Temperatures across the wing and fuselage then rose, and sensors of pressure, temperature and other parameters failed.

The investigation is still in its evidence-gathering phase, says Waldock, and firm conclusions are weeks, if not months, away. ■

## DNA study deepens rift over Iceland's genetic heritage

**Alison Abbott**

The row between Icelandic company deCODE Genetics and its opponents has been rekindled by an analysis of European DNA sequences.

Based in Reykjavik, deCODE has claimed that Icelanders are relatively genetically homogeneous, and that analysis of the population's sequence data and medical history should therefore aid the hunt for disease-related genes. To those ends, the Icelandic parliament granted deCODE exclusive access to the medical records of the country's inhabitants. Opponents such as lobby group Mannvernd object to the commercial use of Iceland's medical records.

The claims for Iceland's homogeneity made by deCODE are based on results published over the past few years. Researchers at the company and at the University of Oxford, UK, compared the sequences of mitochondrial DNA samples from about 400 Icelanders with those from other nationalities stored in various sequence databases (A. Helgason *et al.* *Am. J. Hum. Genet.* **66**, 999–1016; 2000).

Einar Árnason, a geneticist at the University of Iceland in Reykjavik and a member of Mannvernd's board, now says that deCODE's researchers were misled by extensive errors in some of the databases they used in their analysis. He also questions a few of the assumptions behind the technique the company used to compare the sequences.

Árnason re-analysed available data, but avoided using databases. He compared mitochondrial DNA sequences generated by his research group and by deCODE, and compared them with data in research papers on 26 different European populations. His results

suggest that several characteristics of the mitochondrial DNA sequences vary as much in Iceland as they do in the rest of Europe (E. Árnason *Ann. Hum. Genet.* **67**, 5–16; 2003).

Scientists at deCODE do not dispute Árnason's sequence data, but say that he has misinterpreted them. The firm's chief executive, Kári Stefánsson, says that the database errors have only a "limited" effect on Iceland's position in the hierarchy of European genetic homogeneity. The company adds that Iceland's extensive genealogical data will aid the hunt for disease genes, whatever the genetic variation within the country's population proves to be. ■



**All mixed up:** Icelanders may be as genetically heterogeneous as the rest of Europe.

B. JEREMY/COREIS SYGMA

## Cancer fears cast doubts on future of gene therapy

**Erika Check, Washington**

A once-promising gene-therapy treatment should be used only as a last resort, advisers to the US National Institutes of Health (NIH) said this week.

Members of the NIH's Recombinant DNA Advisory Committee, meeting in Bethesda, Maryland, on 10 February, said that patients suffering from severe combined immunodeficiency disease (SCID) should be treated by gene therapy only if they fail to respond to all other treatments. SCID, which disrupts the development of the immune system, is the only disease so far to have been cured by gene therapy.

SCID trials in the United States were

suspended last year by the Food and Drug Administration (FDA) after a child in the world's most advanced study, led by Alain Fischer of the Necker Hospital for Sick Children in Paris, developed leukaemia. The agency halted 24 similar gene-therapy trials when a second child in Fischer's trial was diagnosed with leukaemia in December (see *Nature* **421**, 305; 2003).

In both cases, the retrovirus used to deliver the corrective gene to the patient inserted itself into a stretch of DNA in or near a gene called *LMO2*, which can cause childhood leukaemia. Since then, Christof von Kalle, a molecular geneticist at the Cincinnati Children's Hospital, has found

that a similar insertion has occurred in a third child in Fischer's study, although this patient has not developed leukaemia.

The committee's advice to limit the use of SCID gene therapy is expected to influence the FDA's final decision on the US trials. The agency will consider whether to halt the SCID trials permanently when its advisory committee meets on 28 February.

Scientists say that this meeting is crucial for the future of gene therapy. "People are very worried that these cases will shut down the entire field of gene transfer," says Diane Wara, a paediatrician at the University of California, San Francisco, who sits on the NIH committee. ■



# London gears up for road congestion charge

David Adam

The streets of London were once said to be paved with gold, but modern-day prospectors are more likely to find them filled with traffic. In a bid to lose London's tag as one of the most congested places in Europe, its local authority is about to start charging people who want to drive into the city centre.

Whether or not the plan succeeds, it will provide the most important test so far of the scientific models used to design such schemes, and transport researchers are itching to see what happens.

"London is essentially being turned into a big laboratory," says David Milne, a traffic researcher at the University of Leeds. "Road pricing schemes have been kicked around in modelling terms in the United Kingdom since the 1960s, but we haven't any real evidence of how people respond."

From 17 February, people wishing to drive into central London during the day will have to pay a £5 (US\$8) fee. The move is intended to force an extra 20,000 daily commuters onto public transport. Cameras will register the number-plate of every vehicle entering a 13-square-kilometre central zone.

If the models prove accurate, city-centre congestion should fall by up to 30%, say those running the £200-million scheme. London is not the first city to introduce such a toll,

but transport experts say that the scheme, which is run by Transport for London, is the largest and most significant of its kind.

Traffic flow in the models is estimated by considering the amount of road space, and how many people want to use that space. If the models have a weak spot, it is likely to be the latter estimate, say transport researchers. Commuters have been interviewed about how they will react to the charge, and the success of the scheme could rest on whether people behave as they said they would.

"Theory and reality can be miles apart," warns a member of the working group that designed the scheme who did not want to be identified. Commuters will be swayed by the actions of those around them, for example, and emptier roads could even tempt bus and train users to get back behind the wheel instead.

Traffic flow around the edges of the charging zone is another area of uncertainty. Traffic levels in these areas could rise as drivers who don't want to pay the charge spill onto other routes. Milne says that traffic could increase as far away as the M25, the already-packed motorway that circles London some 30 kilometres out. But Transport for London points out that the working group's models predict that traffic in these areas will increase by 6% at most, and that



London's congestion charge is intended to force 20,000 commuters to switch to public transport.

EPA/PA

levels there are lower anyway.

Transport for London has already announced a programme to monitor both of these areas, along with other effects of the congestion charge. The first results, based on follow-up interviews, vehicle-counting and measurements of average journey times, will be published in six months' time, although natural fluctuations in traffic levels mean that it may be years before solid conclusions can be drawn.

♦ [www.cclondon.com](http://www.cclondon.com)

# Long-lost wave report sinks asteroid impact theory

Rex Dalton, San Diego

A forgotten report on ocean waves generated by nuclear explosions has surfaced, forcing researchers to rethink their theories on asteroid impacts.

The *Handbook of Explosion-Generated Water Waves* was completed in 1968 at the request of the US Office of Naval Research by William Van Dorn of the Scripps Institution of Oceanography in La Jolla,

California. But it was not entered into an academic library catalogue until March 2002, when Scripps' library did so as part of a project to record old technical reports.

Since the report was written, planetary scientists have investigated the potential effect of small asteroids — particularly those with diameters of 100–500 metres — landing in the ocean, and have concluded that the resulting tsunamis could devastate regions over 20 kilometres inland.

But new analyses of Van Dorn's highly technical report on waves produced by nuclear explosions, which takes account of factors such as the absorption of wave energy by continental coastal shelves, indicate that tsunami damage would be limited to immediate coastal areas.

Jay Melosh, a planetary scientist at the University of Arizona in Tucson, tracked down the report last year. He plans to talk at the Lunar and Planetary Science Conference in League City, Texas, next month about his own analyses that predict limited inland damage from asteroid-generated waves. "It appears the defence community has already determined that explosion-generated waves

are neither a serious threat nor a promising weapon," writes Melosh in his abstract.

Such risk assessments are extremely important as governments consider the costs of telescopes or satellites needed to conduct asteroid surveys. Scientists agree that the impact of an asteroid larger than 1 km in diameter would be catastrophic, but there is much debate over the risk from smaller ones. If Earth is under threat only from larger asteroids, then equipment and research costs will be much lower than if scientists also have to look out for numerous smaller objects.

The US government is funding the development of the Panoramic Survey Telescope and Rapid Response System and the Large Synoptic Survey Telescope, whose brief includes surveys of asteroids of various sizes. The latter telescope is expected to cost around \$200 million.

A NASA task force is to complete a report in the spring on the risk from smaller asteroids. "It is fortunate this very important report has come out now," says Steve Chesley, a planetary scientist at NASA's Jet Propulsion Laboratory in Pasadena, California, and a member of the task force.



# Hints of age bias spur calls for grant reforms

David Cyranoski and Keiko Kandachi, Tokyo

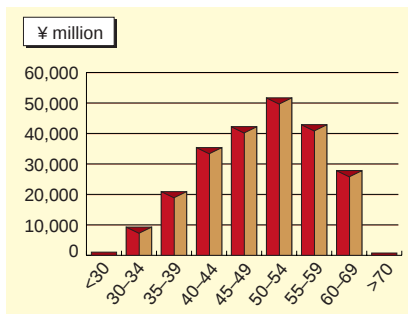
The older you are, the more funding you get, according to figures released earlier this year by Japan's highest science and technology body, the Council for Science and Technology Policy (CSTP).

Now the CSTP has set up a committee to debate what measures, if any, should be taken to redress the balance and increase financial autonomy for young researchers. It is scheduled to report back in March.

The figures were compiled from all seven of Japan's grant-giving agencies for the fiscal year 2001; the first time such a complete survey has been done. They reveal a marked imbalance, showing that 71% of the money goes to researchers aged 45 or over, with those over 50 getting more than half of all grant funding.

The issue isn't confined to Japan: figures released last year by the US National Institutes of Health (NIH) revealed that principal investigators aged 40 and under got only 16% of the 6,961 competitive research grants that it awarded in 2001.

For some, this pattern is out of sync with the age at which researchers are most productive. "We need to shift more funds to those younger people," says Masahiko Aoki, an economist at Stanford University and a special adviser to the CSTP's review committee. He points to data on the CSTP's website



Uneven: grant funding in Japan by age for 2001.

that suggest that most Nobel laureates do their prizewinning research in their thirties.

But Japan's bureaucrats are not convinced. An official at the education ministry says that thanks to programmes set up five years ago, 28% of the ministry's grants for 2001 went to researchers under 40. Even officials at the health ministry, where only 2.4% of last year's 1,400 grants went to researchers under 36, were unconcerned. The ministry's grants are "mission-oriented" and not intended "to give young people funds to be creative with", says one official.

Hiroo Imura, chairman of the CSTP committee and the former president of Kyoto University, sees the real issue as whether a small, established group of older scientists

gets too much funding. "We think it's important to avoid giving several government grants to a handful of researchers," he says.

Immunologist Kimishige Ishizaka, who is another adviser on the CSTP committee and was formerly at the La Jolla Institute for Allergy and Immunology in California, blames both universities and funding agencies for this effect. Neither steps in to ensure that researchers are not overextending themselves or duplicating research, he says.

But like Aoki, Ishizaka thinks that the real problem is that young researchers cannot be independent, and he blames Japan's brief application forms, which usually describe a research plan in just two pages. "You can't tell what's unique or how the proposed project compares to others," he says. "It's impossible to evaluate on this basis." So decisions are based mainly on past records, which favours older researchers, Ishizaka explains.

Although figures show that Japanese under 40 are relatively successful at getting grants (27% compared with the NIH's 16.3%), the average grants are only 33–50% as large as those received by their older colleagues. Ishizaka says that Japan should break down some of its larger projects and spread out the money because the grants awarded to young researchers are often too small to support truly independent research. ■

SOURCE: CSTP

# Civil war leaves Ivory Coast research in tatters

Marieke Degen, Munich

The civil war raging in the Ivory Coast is counting scientific research among its many victims. Fighting and looting in what was one of Africa's wealthier nations have left international projects on health, agriculture and the environment in disarray.

When conflict between rebel and government troops erupted five months ago, foreign researchers working in disputed regions had to leave their scientific stations for their own safety. For some, the work that they left behind may never be recovered.

Last September, a team led by Eduard Linsenmair, an ecologist at the University of Würzburg, Germany, had to abandon its biodiversity studies in the Comoé National Park. The group has since heard that government troops have impounded the research station's cars and motorcycles.

"We learned that all of the buildings have been broken into and looted," Linsenmair says. Some of his projects will now be relocated to Benin, Burkina Faso and Ghana, but a lot of the data have been lost, he says.

Others have been a little more fortunate. In September, when rebels took over Bouaké,

the Ivory Coast's second largest city, the 30 international researchers at the West Africa Rice Development Association (WARDA) were evacuated. They left behind one of Africa's most important rice gene banks. But some of the team returned on 17 December and recovered some 6,000 samples of recently developed rice varieties, says WARDA's director, Kanayo Nwanzé. The association is now moving some of its operations to Mali.

Yasmin Möbius and Tobias Deschner from the Max Planck Institute for Evolutionary Anthropology in Leipzig also returned to their project in December. But they found that irreversible damage had been done.

Möbius and Deschner had been studying chimpanzee behaviour in the Tai National Park. Because the chimps are used to humans, they are easy prey for poachers. Soon after the researchers left the station in September, three female and three young chimps, including rare twins, were killed for food.

"We returned voluntarily," says Möbius. "It was important for us to ward off further attacks on the animals."

The situation in the national park stayed calm until 3 January, when fighting broke out

near the Liberian border. The researchers returned to Abidjan, but left for Germany after fighting in the capital intensified.

A peace plan to end the war was discussed in Paris on 24 January, but has been rejected by supporters of President Laurent Gbagbo.

"The war is a catastrophe," says Linsenmair. "All we can do is wait and hope it will end." ■







Road to ruin or road to riches? Indian farmers taking their cotton crop to market last month.

## India debates results of its first transgenic cotton crop

K. S. Jayaraman, New Delhi

India has just collected its first commercial harvest of transgenic cotton. But even as the last bolls were being picked, arguments were beginning over the crop's success.

About 50,000 farmers planted transgenic cotton in India last year, after the government licensed commercial use of the crop in March. The plants carry a gene for producing a toxin from the bacterium *Bacillus thuringiensis* (Bt), making them resistant to the bollworm caterpillars that ruin up to half the crop each year.

Assessments are being muddled by the lack of a large-scale, independent survey. The season was hailed as a success by the company that sells the seeds, Mahyco Monsanto Biotech (MMB) — a joint venture between Maharashtra Hybrid Seed Company in Jalna and Monsanto in St Louis, Missouri. Yields were up and pesticide use down, the firm claims. Environmental groups, on the other hand, report the opposite effects.

An analysis of the field trials that preceded the licensing indicates that transgenic cotton may benefit Indian agriculture, however. Agricultural economists Matin Qaim of the University of Bonn in Germany and David Zilberman of the University of California, Berkeley, examined controlled trials of Bt cotton grown in 157 farms across three Indian states. Their results, published last week (M. Qaim and D. Zilberman, *Science* **299**, 900–902; 2003), show that transgenic varieties gave 80% higher yields and allowed farmers to cut pesticide use by 70% compared with normal cotton. “The farmers were very positive,” says Qaim.

Bt cotton reduces total cost when there is a heavy bollworm infestation, says

Ebrahimi Abubacker Siddiq, former deputy director of crop science at the Indian Council of Agricultural Research, and chair of the committee that monitored the trials and the commercial planting. But he points out that the government's analysis of the 2001 trial puts the increase in yield at 20–60%. “Claiming increased yields of 80% will raise farmers' expectations,” he cautions.

A number of issues remain to be resolved, warn researchers, including the possibility that bollworms may develop resistance to the insecticide produced by the plants. The government requires farmers using Bt cotton to grow ordinary cotton on 20% of their land to delay the spread of resistance. But Devinder Sharma of the Forum for Biotechnology and Food Security, an environmental group based in New Delhi, says that many farmers failed to comply with this condition.

Absence of a mechanism to assess long-term effects is another problem, adds Siddiq. “Hardly anything is known about how long the leftover biomass of transgenic crops would remain in the soil and what kind of environmental impact it would have,” he says.

Such issues are being studied by groups on all sides of the debate, while state governments and agricultural researchers push to introduce more transgenic crops to India. Insect-resistant varieties of rice, chickpea and a pulse called pigeon pea are awaiting large-scale trials. The department of agriculture in Andhra Pradesh is hoping to introduce insect-resistant varieties of sorghum and peanuts within two years. In December the National Academy of Agricultural Sciences endorsed the introduction of salt-tolerant rice, now undergoing trials. ■

Additional reporting by John Whitfield, London.

## Social scientists call for abolition of dishonesty committee

Alison Abbott

Following its controversial ruling on political scientist Bjørn Lomborg's book *The Skeptical Environmentalist*, the Danish Committees on Scientific Dishonesty (DCSD) is now the subject of debate itself.

The DCSD has quietly ruled on an average of one case of alleged misconduct per year since it was established in 1992. But many researchers felt it overstepped the mark by investigating Lomborg's book, which paints an unusually rosy picture of the global environment.

In a report published on 6 January, the DCSD said that the book was “objectively speaking, deemed to fall within the concept of scientific dishonesty” (see *Nature* **421**, 195 & 201; 2003). The decision triggered debate in the Danish parliament and newspapers. Last week science minister Helge Sander asked the Danish Research Agency to set up an independent working group to examine the regulatory basis and procedures of the DCSD.

The affair has split Denmark's academic communities. Given Lomborg's background, many social scientists think that his book should not be judged by criteria used to assess dishonesty in the natural and medical sciences.

Jørn Henrik Petersen, a social historian at the University of Southern Denmark in Odense, says that selection of information to develop a theory — anathema to natural scientists — is an integral part of many social sciences. “It is out of the question to argue that any selection can be completely objective,” he says. Petersen is one of many social scientists who say that the DCSD should be disbanded.

Some 600 natural and medical scientists in Denmark have signed a petition in support of the DCSD that has been presented to the Danish Research Agency. “We don't object to the procedures being re-examined, but it is expedient for any society to have this kind of committee,” says Jens Rehfeld, a health researcher at the Rigshospitalet in Copenhagen.

The working group, including social and natural scientists and journalists, will be chaired by Mogens Pedersen, a political scientist at the University of Southern Denmark. Key issues will be whether the definition of ‘scientific dishonesty’ should be changed, what kinds of work should be included in the DCSD's remit, and how the results of any deliberation should be presented to the public. It is expected to report before the summer. ■

## Spending spree gives German physics a high-energy boost

**Munich** A free-electron X-ray laser and a heavy-ion and antiproton accelerator are among four projects given the green light last week by the German government.

Announcing the E1.6-billion (US\$1.7-billion) funding package on 5 February, the education and research minister Edelgard Bulmahn also pledged money for a new E100-million research aircraft for Earth and atmospheric research. The spending spree includes E25 million for a laboratory near Dresden that will use powerful magnetic fields for materials science and studies of condensed matter.

The E673-million free-electron X-ray laser will be situated at the German Electron Synchrotron in Hamburg. Free-electron lasers use electromagnetic radiation from particle accelerators to generate intense X-ray radiation with ultrashort wavelengths. Chemists, structural biologists, materials scientists and plasma physicists use them to study molecules during chemical reactions.

The facility, which will be jointly financed by several European countries, is scheduled to become operational in 2011. The antiproton accelerator, which will generate high-intensity ion beams and comes with a price tag of E675 million, will be built at the Heavy Ion Research Centre in Darmstadt.

## Neuroscientists cruise for research-centre funding

**San Diego** Caribbean cruises are a rare treat for researchers, so an invitation from billionaire and Microsoft co-founder Paul Allen for a dozen neuroscientists to spend time aboard a yacht set tongues wagging.

Nature has learned that the cruise, held last spring, was one of a series of secret meetings between leading neuroscientists and Allen, who is planning to fund a new research institute. Boat trips apart, the plans are being kept under close guard. Officials at the Paul G. Allen Foundations in Seattle, Washington, declined to discuss the issue, and scientists consulted on the project have signed confidentiality agreements.

The institute is expected to be set up in the Seattle area. One of its first projects is likely to be a molecular map of the mammalian brain, to show gene-expression patterns in the mouse nervous system.

## Indian cricketers team up to hit polio for six

**New Delhi** Public-health officials in India are hoping to use the current Cricket World Cup to defeat childhood polio in the country. The largest ever mass immunization against the



**Signing up:** Indian cricket captain Sourav Ganguly puts his name to a campaign to beat polio.

disease began on 9 February — the first day of the tournament in southern Africa. By the time the Indian team returns home, the officials hope to have protected all of the country's 165 million children under 5 years of age.

"We need a massive social mobilization effort," says Savita Varde-Naqvi, a spokesperson for children's charity UNICEF in New Delhi. "And if cricket captain Sourav Ganguly tells kids to get vaccinated against polio, they will." The team has recently promoted polio vaccinations through adverts, competitions and events.

The move marks the start of a large-scale attempt to bring the virus, which causes paralysis and death in children, back under control. The global campaign to eradicate polio reduced the number of new cases worldwide in 2001 to just 483, but last year there were 1,556, with 85% of them in India.

## NIH aims to characterize embryonic stem-cell lines

**Washington** Tired of comparing apples with oranges? Researchers working on human embryonic stem cells are — the cell types available to them are poorly defined, behave differently to each other in the lab, and make results difficult to compare. To address the problem, the National Institutes of Health (NIH) has set up a unit to characterize the stem-cell lines that are currently available to federally funded scientists.

The project, to be run by Ron McKay of the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, aims to compare the different cells' surface molecules and the cell lines' patterns of gene expression. The project could grow to answer bigger questions, such as whether the cell lines can generate every type of cell in the body. The NIH will post its results on the Internet.

"A uniform evaluation of all available lines by a single lab would be enormously valuable," says George Daley of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts.

## Rare book collection faces final chapter

**Caracas** An unrivalled Latin American collection of books and journals is facing ruin as a result of Venezuela's political crisis.

The Marcel Roche Library at the Venezuelan Institute for Scientific Research (IVIC) in Caracas has had its promised US\$3.2 million of funding withdrawn by the government, library officials say. The money was earmarked for purchasing new journals and maintaining the current collection.

IVIC officials have issued an international appeal for funds or assistance. The collection includes many complete series of journals, including *Philosophical Transactions of the Royal Society* from its first issue in 1665. In 1996 it was named regional library for Latin America and the Caribbean by UNESCO.

## University grilled over bacon chops' transgenic link

**San Francisco** This little piggy went to market, and the scientists who created its transgenic parents are facing some awkward questions. The US Food and Drug Administration (FDA) has launched a probe into whether genetically engineered pigs from research labs have crept into the food supply.

It emerged last week that nearly 400 piglets descended from genetically modified animals at the University of Illinois have been sold to a local livestock trader since 2001, and many could have subsequently appeared in bacon sandwiches. The FDA wants to know whether any were carrying transgenes present in the adult pigs. One is a bovine gene that stimulates lactation, the second is a synthetic gene for insulin-like growth factor, intended to improve digestion.

The university says that the transgenes were not passed on to the piglets, but FDA officials have not confirmed this. Even if the piglets were unaffected, the university could still be fined for selling them without permission.



**Pig issue:** offspring from genetically modified pigs may have entered the food chain.



# On wings and a prayer

In the wake of Columbia's loss, NASA's efforts to replace its ageing shuttle fleet are coming under fresh scrutiny. Geoff Brumfiel uncovers a tale of high hopes, false starts and immense technical hurdles.

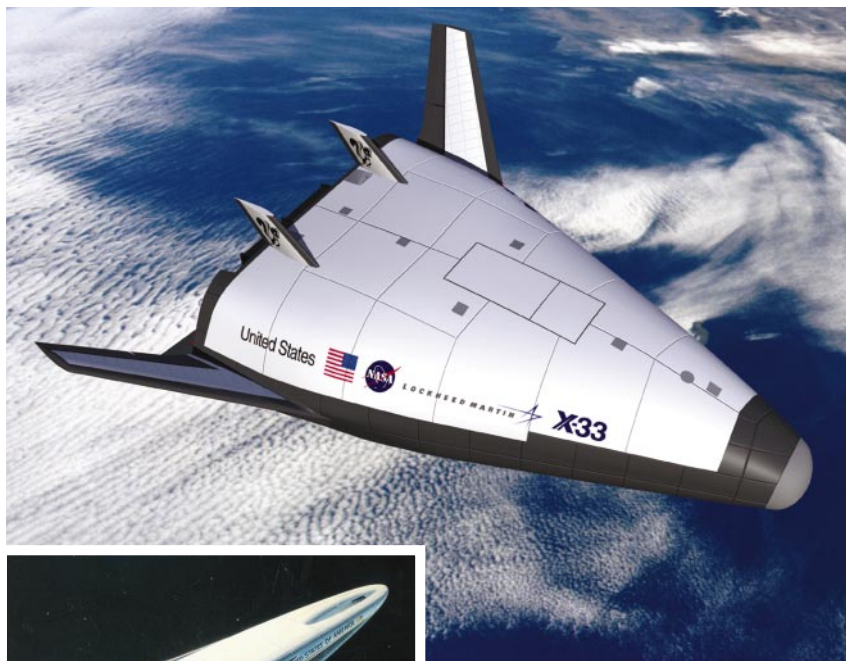
In 1986, just weeks after the space shuttle Challenger exploded, President Ronald Reagan announced plans to develop a successor to the craft. This "new Orient Express", as he called it, would speed from Washington to Tokyo in just two hours, flying at the edge of space.

Officially known as the National Aerospace Plane, the craft was to fly at 25 times the speed of sound, moving effortlessly in and out of orbit. It would be free of the bulky external fuel tanks and boosters of its shuttle predecessor. But, like so many concept vehicles advocated over the two decades since the shuttles began flying, it fell victim to formidable technical obstacles. In 1994, the project was cancelled, writing off some US\$3 billion of research.

NASA's current plans, announced shortly before the Columbia disaster, are much less ambitious. Last November, after halving the agency's five-year, roughly \$5-billion budget for new space-launch technologies, President George W. Bush's administration unveiled a two-pronged approach to replacing the shuttles. The first is to develop an Orbital Space Plane within a decade that, unlike Reagan's 'Orient Express', will be wedded to conventional launch technologies. The second is an open-ended study of advanced concepts which has no firm deadline for producing a new system.

Following Columbia's loss, those plans may shift once more. But it is clear that we're not going to see anything along the lines of the National Aerospace Plane in the foreseeable future. Ideally, spacecraft would be more like conventional aeroplanes. They would take off, fly to orbit and land in a single stage, needing only to be refuelled between flights. They would carry fuel in internal tanks, manoeuvre gracefully both in the atmosphere and in space, and be able to withstand the intense heat of re-entry. But meeting all of these requirements has so far proved beyond the capabilities of the world's leading aerospace engineers.

Given what happened to Columbia, manoeuvrability within the atmosphere and surviving re-entry are clearly key issues. The space shuttle looks rather like a brick with wings, and for a good reason. Its blunt shape



**It'll never fly: the ambitious National Aerospace Plane (inset) and the X-33 were each hailed as the future for space travel. But spiralling costs and insurmountable technical problems eventually scuppered both designs.**

creates a shock wave as it hurtles back into the atmosphere, which helps to protect the craft from the hot ionized gases that form around it. If it looked and could fly like a high-performance jet, the shuttle's leading edges would heat far beyond the roughly 1,600 °C they currently experience. Scientists at NASA's Ames Research Center in Moffett Field, California, are working on ceramic materials, based on composites of hafnium or zirconium diborides with silicon carbide, that might help a more manoeuvrable vehicle beat the heat of re-entry. These materials can survive temperatures of nearly 2,800 °C, but are currently brittle and bulky, and more research will be needed before they can be incorporated into designs for sleek spaceplanes.

By far the biggest obstacle for any spacecraft, however, is getting into orbit in the first place. To escape the pull of Earth's gravity, a

craft must accelerate to more than 40,000 kilometres per hour. Launching a 22-tonne payload to these speeds currently requires a rocket weighing nearly 1,000 tonnes, which consists mostly of liquid hydrogen and oxygen fuels and the tanks that hold them. Today's rockets shed their emptied tanks as they climb into space, but many aerospace engineers consider this 'stage' strategy a wasteful and inelegant solution to the launch problem.

## Weighty issues

Over the decades, engineers have racked their brains to overcome the weight issue and move to a single-stage vehicle. As fuel makes up most of a rocket's weight, the simplest solution would be to find a lightweight, super-efficient propellant, but so far, none has emerged. "When Captain Kirk decides to send us some dilithium crystals, that will change," says T. K. Mattingly, a

LOCKHEED MARTIN

FAS

former Apollo and shuttle astronaut who has devoted the past 20 years to developing advanced launch-vehicle concepts. "But until then we have to live with chemistry."

The National Aerospace Plane would have used another trick to do away with much of its bulky fuel. Rather than carrying large amounts of liquid oxygen, the craft would have scooped it as a gas out of the atmosphere at hypersonic speeds, compressed it, and combusted it with on-board liquid hydrogen in a special chamber. After accelerating to many times the speed of sound, the plane would head upwards and zoom into orbit.

This novel engine design is known as a scramjet, and involves some major technical challenges. Because scramjets can only work in the thin atmosphere of very high altitudes, and at speeds greater than five times that of sound, such planes must incorporate more conventional engine technology for take-off and landing. And, as there is no oxygen in space, these planes would still need to carry some liquid oxygen on board to use once they are in orbit.

These complex technical needs proved to be too much for the National Aerospace Plane. "The programme lost political and financial backing because there wasn't enough confidence that you could make it work," says Vance Brand, who worked on the plane and is now acting director of aerospace projects at NASA's Dryden Flight Research Center in Edwards, California.

### Lightening the load

Besides cutting the amount of fuel being carried, it may also be possible to reduce the weight of the spacecraft itself. This became NASA's leading strategy after the National Aerospace Plane's demise. In 1996, the agency selected contractor Lockheed Martin of Denver, Colorado, to help design a lightweight craft that could get into orbit in a single stage. The company turned the task over to its famous 'Skunk Works' design team in Palmdale, California. The result was the X-33, a wedge-shaped experimental aircraft that NASA hoped would prove the feasibility of a self-contained spaceship.

"Lockheed's approach was very bold," says Andrew Butrica, a former NASA historian who is now writing a book on the X-33 programme. The craft used graphite-epoxy composites that were a fraction of the weight of the aluminium conventionally used for spacecraft construction. But these high-tech materials were poorly understood and proved difficult to work with. "The thing that ultimately failed was the hydrogen tank," says John Paulson, who heads the Vehicle Analysis Branch at NASA's Langley Research Center in Hampton, Virginia. Among other problems, the craft's tanks had trouble dealing with the thermal stresses of holding fuels that were a few degrees above

absolute zero. In a 1999 test filling, the walls of the hydrogen tank failed. After spending more than \$1 billion on the project, NASA and Lockheed Martin bailed out in 2001.

Today, NASA's programme to replace the shuttle is much more conservative. Rather than betting heavily on advanced engines or lightweight composite materials, the agency plans to build a small craft called the Orbital Space Plane. The working designs are not revolutionary — in fact, one leading contender, put forward by Orbital Sciences of Dulles, Virginia, and Northrop Grumman of El Segundo, California, is based on a Soviet concept spacecraft called BOR-4, which was first spied by the West in the early 1980s. The Orbital Space Plane will be strapped to the top of a conventional, disposable rocket booster and shot into space. The plan is for it to be used initially as a 'lifeboat' for the International Space Station, and later as a taxi service for rotating the station's crews.

### One small step

Paulson admits that the Orbital Space Plane is "not going to be a tremendous leap". But he stresses that practical considerations make it the next logical step. Mattingly concurs, noting that rocket pioneer Wernher von Braun was once asked what he thought was the best launch vehicle. "He answered: 'Whatever is affordable,'" says Mattingly.

Meanwhile, NASA and the Department of Defense continue to devote modest funding to concepts such as scramjets and advanced materials that might eventually

lead to a single-stage reusable space vehicle. Many experts believe that this may come from an intermediate, two-stage system in which a scramjet-powered craft launches a rocket from the upper atmosphere.

The big question is whether the political will exists to bring these projects to fruition. The space shuttles were originally intended to launch scientific, military and commercial satellites — indeed, in the early days of the shuttle programme, US government satellites had to be launched on the craft by congressional order. But following the 1986 Challenger accident, those rules were abandoned. Today, satellite launches depend on expendable rockets. And in the light of the current slump in the airline industry, and the continuing risks of space travel, Reagan's vision of travellers zooming from Washington to Tokyo now looks more than a little starry-eyed. "I'm very pessimistic in the near term," says Butrica, "and I'm fearful that Columbia has dealt a near deathblow to the manned space programme."

Mattingly is certain that the technical obstacles to creating a fully reusable spaceplane can be overcome. But he feels that real progress will come only if enthusiasts have a clear goal at which they can drive their technology programmes. "We need to stop and ask ourselves: 'What are we trying to do?'" he says.

Geoff Brumfiel is *Nature's* physical sciences correspondent in Washington.

NASA's Space Launch Initiative  
[www.slinews.org](http://www.slinews.org)



Looks familiar: NASA's current vision for the next generation of space vehicle is the Orbital Space Plane. But one leading design (above) owes more than a small debt to Soviet technology of the 1980s in the shape of BOR-4 (inset).



# Altered states

Changes to the genome that don't affect DNA sequence may help to explain differences between genetically identical twins. Might these 'epigenetic' phenomena also underlie common diseases? Carina Dennis investigates.

**M**arie and Jo — not their real names — are twins, genetically identical and raised in a happy home. Growing up, both enjoyed sport and art, were equally good at school, and appeared similar in almost every respect. But as adults, their lives and personalities diverged: in her early twenties, following episodes of hallucinations and delusions, Marie was diagnosed with schizophrenia.

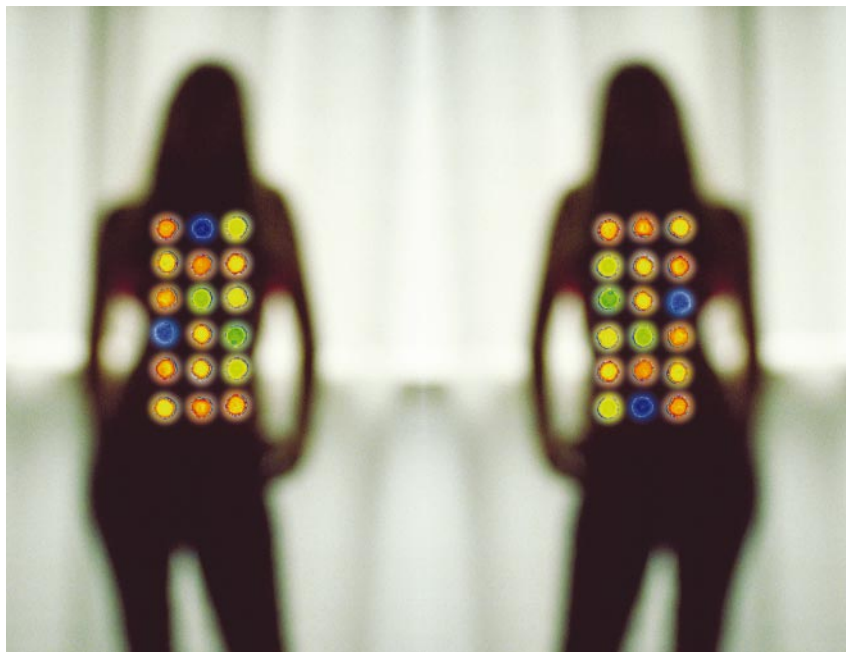
Such examples have long baffled geneticists. Despite sharing the same DNA and often the same environment, 'identical' twins can sometimes show striking differences. Now some researchers are beginning to investigate whether subtle modifications to the genome that don't alter its DNA sequence, known as epigenetic changes, may provide the answer. In doing so, they hope to shed light on the mysterious roots of common diseases — such as schizophrenia and diabetes — that burden society as a whole.

The idea that epigenetics underpins many of the world's health scourges is still highly speculative. "Most geneticists believe that the essence of all human disease is related to DNA-sequence variation," says Arturas Petronis, a psychiatrist at the University of Toronto in Canada. But with the genomics revolution having yet to yield the hoped-for avalanche of genes that confer susceptibility to common diseases, Petronis is not alone in believing that it's time to revisit the problem under the spotlight of epigenetics.

Each of our cells carries the genes for making all the building blocks of the body, but only some of them are active. Epigenetic modifications act like switches, helping to control gene activity so that only those that are required in a particular cell are actually turned on. They constitute a 'memory' of gene activity that can be passed on each time a cell divides, ensuring that liver cells beget more liver cells, and so on.

The best-known epigenetic signal is DNA methylation, which tags cytosine, one of the four chemical bases that make up the genetic code, with a methyl group. Although not a hard-and-fast rule, DNA methylation is generally associated with silencing of gene expression, whereas active genes are usually unmethylated.

Chromatin remodelling is another



Spot the difference: colour-coded epigenetic microarrays could show why 'identical' twins aren't the same.

important epigenetic mechanism. Inside the nucleus, strands of DNA wrap around proteins called histones, and then coil up again into a densely packed structure called chromatin. Chemical modification of the protruding histone tails — by the addition of acetyl, methyl or phosphate groups — can alter chromatin structure, which in turn influences the activity of adjacent genes. There are some predictable patterns — for example, genes associated with acetylated histones are usually switched on. But the consequences of the many different combinations of histone modifications — known as the histone code — have yet to be fully deciphered.

Together, these mechanisms modulate gene expression throughout the genome, and underpin several unusual phenomena, including the shutting down of one copy of the X chromosome that occurs in female mammals, and parental 'imprinting' — in which a gene's activity depends on whether it is inherited from the mother or the father.

Over the past few years, some researchers studying rare diseases have found themselves delving into epigenetics. In some cases, they

stumbled into the subject from conventional genetics, discovering that the mutated gene for a rare condition exerts its effect by influencing epigenetic modifications elsewhere in the genome, for instance. Children born with mutations in a gene called *MECP2*, which influences chromatin remodelling, gradually lose their ability to speak and walk<sup>1</sup>. People with a mutant form of a gene for an important DNA-methylating enzyme suffer from immune-system deficiencies and facial deformities<sup>2</sup>. And mutations in a gene called *ATRX* cause mental retardation, urogenital abnormalities and a form of anaemia — apparently the result of changes in both chromatin structure and DNA methylation<sup>3</sup>.

Among oncologists, the idea that epigenetic changes might be linked to disease is already well established. "Epigenetic changes are terribly important in cancer," says Andrew Feinberg, a geneticist at Johns Hopkins University in Baltimore, Maryland. Two decades ago, Feinberg and his colleague Bert Vogelstein were among the first to show that cancer cells may exhibit unusual patterns of DNA methylation<sup>4</sup>. Since then, similar reports of epigenetic





A methylating enzyme (white) binds to its target site (red) on DNA; the methyl donor is shown in green.

abnormalities in cancer have proliferated<sup>5</sup>.

Yet researchers who are interested in other common diseases — such as diabetes, obesity, heart disease and a host of psychiatric disorders — have shown relatively little interest in epigenetics. Whereas tumours arise from aberrant cells, making it easy to imagine how an epigenetic mix-up might be involved, the origins of other conditions are more complex, which made the idea seem less plausible for these diseases. “It’s not been on people’s radar screens,” says Feinberg.

### Disease by descent

Besides, many common diseases seem to be inherited within families, which is incompatible with the long-standing idea that the epigenetic slate is wiped clean in the embryo shortly after fertilization, except for parentally imprinted genes. Only now is evidence emerging that other epigenetic changes can be inherited (see ‘Keep it in the family’, right).

The lure of genomics has also been instrumental in keeping medical geneticists focused on DNA sequence rather than what goes on around it. One consequence of the Human Genome Project has been the mapping of more than a million single-nucleotide polymorphisms (SNPs) — positions in the genome at which the sequence varies between individuals by a single letter in the genetic code. By seeing which SNPs are inherited alongside common, complex diseases, it should be possible to pin down the locations of genes that make some people particularly susceptible to certain conditions. But SNP-based approaches have so far yielded a disappointing harvest, which has led some researchers to wonder whether sequence variation is just part of the story. “The role of epigenetics in common diseases has been underestimated —

it’s the new frontier,” claims Feinberg.

That frontier is still a fairly lonely place, but pioneers are starting to explore. “Eight years ago, I didn’t even know what the word ‘epigenetics’ meant,” says Petronis. Today, he feels that some features of complex diseases are easier to explain in terms of epigenetic changes than through conventional genetics. Some conditions, such as multiple sclerosis, fluctuate in severity as they progress. Although the case is far from cut-and-dried, Petronis believes this to be more consistent with shifting epigenetic profiles than with the constant nagging presence of a susceptibility gene<sup>6</sup>.

Factors such as lifestyle and diet undoubtedly influence our susceptibility to disease, and there is mounting speculation that they leave a trail of epigenetic footprints across our genome. Alexander Olek, chief executive officer of Epigenomics, a Berlin-based company that aims to hunt down the epigenetic markers of disease, is hot on the trail of changes associated with diet. Olek hopes that this will lead him to an epigenetic fingerprint for diabetes<sup>7</sup>. Epigenetics may also help to explain why many diseases only appear late in life. “Age is accompanied by changes in DNA methylation,” Olek points out.

While Olek’s firm is launching a wide-ranging effort to study disease epigenetics, Petronis was drawn to the subject by his long-standing quest to understand the most debilitating conditions that haunt the human mind, including bipolar disorder — popularly known as manic depression — and schizophrenia. Many features of these conditions, such as the differences between twins like Marie and Jo, cannot be explained readily by DNA-sequence variation. And although some psychiatric disorders seem to run in families, your chance of succumbing depends

## Keep it in the family

You are what you eat, or so we are told. But how could what your grandmother ate make any difference? That’s the mystery posed by the grandchildren of women who were pregnant during the famine that hit the Netherlands at the end of the Second World War. Not only were the babies born to malnourished women smaller than normal, but the next generation also had unusually low birth weights<sup>13</sup>.

A similarly perplexing finding emerged last November, in a study of the grandchildren of Swedish men born in 1890–1920. From crop records, Gunnar Kaati and his colleagues at the University of Umeå calculated how much the grandparents had eaten just before puberty. They found that the grandchildren of well-fed adolescents had a greater risk of dying from diabetes, whereas those descended from famine survivors were less likely to die of heart disease<sup>14</sup>.

How can this be? It smacks of Lamarckism, the discredited theory that evolution occurs by the inheritance of acquired characteristics. Now some researchers argue that the answer lies in epigenetics. “Changes in diet could activate certain pathways that then leave an imprint that is passed to the next generation,” says Alexander Olek of the Berlin-based company Epigenomics.

It’s a contentious idea — biologists have long thought that the genome is reprogrammed around the time of fertilization. But recently, evidence has emerged that some epigenetic changes survive this process.

At the University of Sydney in Australia, researchers led by Emma Whitelaw have studied a strain of genetically identical mice whose fur ranges in colour from yellow to black, depending on the methylation of a single gene. They found that coat colour tends to pass from mothers to their pups<sup>15</sup>. And now a study of fruitflies is generating a similar stir. Douglas Ruden, of the University of Alabama at Birmingham, and his colleagues found evidence that reductions in the activity of a protein called Hsp90 cause heritable changes in eye structure that are the result of epigenetic changes in chromatin<sup>16</sup>.

If similar phenomena occur in humans, it is possible that epigenetic modifications, and not just mutations in DNA sequence, could lie at the root of inherited diseases.



Hungry hopefuls: a Dutch food queue in 1945.



**DNA detective:** Stephan Beck is aiming to uncover modifications that underlie immune diseases.

in some cases on whether this history is on your mother's or father's side — suggesting, perhaps, that parental imprinting is involved.

Previous genetic studies of schizophrenia and bipolar disorder had seemed to implicate regions of chromosome 22 (refs 8, 9), but the search for a susceptibility gene has drawn a blank. So Petronis is now analysing methylation patterns across the chromosome in DNA from post-mortem samples of brain tissue from affected and unaffected individuals. He then plans to home in on those regions that show clear differences, and look for altered patterns of gene expression.

Attracting more researchers to the epigenetic frontier may require the counterpart of today's high-density SNP maps: a systematic effort to map epigenetic variation across the genome. That is the mission of the Human Epigenome Consortium<sup>10</sup>, which is cataloguing the genomic positions of distinct methylation variants. The consortium, established in December 1999, includes Epigenomics, as well as Britain's Wellcome Trust Sanger Institute near Cambridge, the French National Genotyping Centre near Paris, the German Cancer Research Centre in Heidelberg, the Technical University of Berlin and the Max Planck Institute for Molecular Genetics, also in Berlin.

The consortium is now documenting the position of methylation variants across the major histocompatibility complex (MHC), a region on chromosome 6 that contains roughly 150 active genes, many of which are involved in immune recognition. "We chose this region because it is associated with more human diseases than any other," says Stephan Beck, the Sanger Institute's head of sequencing.

A pilot study involving eight initial tissue types has so far identified 4,500 sites within the MHC at which DNA methylation can occur. Information about these sites, which

will shortly be made freely available to other researchers, should prove especially useful for researchers looking for clues to why some people develop autoimmune diseases.

Such studies will involve screening tissue samples from large numbers of people with the condition, and from unaffected controls, to see if a particular methylation fingerprint is strongly correlated with the disorder. Here, the bottleneck is the current lack of high-throughput technology to allow hundreds or thousands of samples to be screened rapidly.

### On the map

One way of mapping methylation sites is to treat DNA with a chemical called sodium bisulphite, which converts unmethylated cytosine to a base called uracil, but leaves methylated cytosine intact. The altered bases can then be detected using standard molecular-genetic probes. Several groups, including researchers at Epigenomics, are now working to incorporate these methods into DNA microarrays that could allow samples to be screened for methylation at thousands of genomic sites in parallel.

High-throughput methods are also being developed to screen for histone modifications. The idea is to treat tissue samples with chemicals that 'freeze' modified histone proteins and their associated DNA in their living state. Antibodies targeted at a particular histone modification can then be used to fish out fragments of chromatin that bear these changes. Finally, their genomic location can be established by analysing their associated DNA, again using microarrays. Currently, however, attempts to develop such methods are dogged by problems, including antibodies' tendency to cross-react with different histone modifications.

While the technology for high-throughput studies of large samples of the general popula-

tion is being perfected, some researchers are turning to pairs of identical twins in which only one suffers from a disease. With the variable of DNA-sequence variation removed, it should be simpler to spot distinctive epigenetic modifications that correlate with the condition. Emma Whitelaw, an epigeneticist at the University of Sydney in Australia, for instance, has teamed up with Nick Martin of the Queensland Institute of Medical Research in Brisbane, who has co-founded a large twin registry. "We are keen to start with diseases with an unambiguous clinical diagnosis," says Whitelaw.

Petronis, meanwhile, is already studying pairs of twins in which one suffers from schizophrenia, and has found substantial differences in DNA methylation. "Not all of these changes will be related to the disease," he warns. "The quest we are now facing is distinguishing epigenetic background 'noise' from functionally relevant signals."

Establishing cause and effect in disease epigenetics is a major problem, agrees Olek. This quest is further complicated by the fact that different tissues — and even different cells within tissues — may show epigenetic differences. Epigenetic modifications can even vary with the conditions under which cells are grown in the lab. "It's overwhelmingly complex," concludes Whitelaw.

To establish links between an epigenetic modification and a disease, researchers might look to see whether the change precedes the occurrence of symptoms in affected tissues. A stronger case could be made if the epigenetic variant is shown to alter gene expression and so affect cells' biochemical activities in a way that can explain the condition's pathology — which may require new animal models of important human diseases.

But in the long run, will such studies offer hope to patients like Marie? One drug that inhibits DNA methylation, called azacitidine, is already being tested for the treatment of some cancers<sup>11,12</sup>. But it acts across the entire genome. If epigenetic changes are shown to underpin psychoses, diabetes and other diseases, we may need more specific treatments. That will pose a challenge, but not an insurmountable one, argues Feinberg. "I think it will be much easier to treat an epigenetic aberration than a sequence mutation," he says. ■

**Carina Dennis is Nature's Australasian correspondent.**

1. Amir, R. E. *et al.* *Nature Genet.* **23**, 185–188 (1999).
2. Xu, G.-L. *et al.* *Nature* **402**, 187–191 (1999).
3. Gibbons, R. J. *et al.* *Nature Genet.* **24**, 368–371 (2000).
4. Feinberg, A. P. & Vogelstein, B. *Nature* **301**, 89–92 (1983).
5. Feinberg, A. P. *Proc. Natl. Acad. Sci. USA* **98**, 392–394 (2001).
6. Petronis, A. *Trends Genet.* **17**, 142–146 (2001).
7. Maier, S. & Olek, A. *J. Nutr. Sci.* **132**, 2440–2443 (2002).
8. DeLisi, L. E. *et al.* *Am. J. Psychiatry* **159**, 803–812 (2002).
9. Kelsoe, J. R. *et al.* *Proc. Natl. Acad. Sci. USA* **98**, 585–590 (2001).
10. Beck, S., Olek, A. & Walter, J. *Nature Biotechnol.* **17**, 1144 (1999).
11. Silverman, L. R. *et al.* *J. Clin. Oncol.* **20**, 2429–2440 (2002).
12. Kornblith, A. B. *et al.* *J. Clin. Oncol.* **20**, 2441–2452 (2002).
13. Lumey, L. H. *Paediatr. Perinat. Epidemiol.* **6**, 240–253 (1992).
14. Kaati, G., Bygren, L. O. & Edvinsson, S. *Eur. J. Hum. Genet.* **10**, 682–688 (2002).
15. Morgan, H. D., Sutherland, H. G. E., Martin, D. I. K. & Whitelaw, E. *Nature Genet.* **23**, 314–318 (1999).
16. Sollars, V. *et al.* *Nature Genet.* **33**, 70–74 (2003).



# Concern about Japan's unclear biotech regulations

To exploit genetics research, agencies need to have clearly defined responsibilities.

**Sir**—The Cartagena Protocol on Biosafety ([www.biodiv.org/biosafety](http://www.biodiv.org/biosafety)), formulated in 2000, is concerned with standardizing the movement of genetically modified organisms (GMOs) across national boundaries. As of January 2003, more than 100 countries have signed and 41 ratified the protocol. Only nine more countries need to ratify the protocol for it to become internationally and universally binding.

Further discussions are needed to reach agreement on the details of each article within the protocol. Article 18, which I discuss here, concerns the handling, transport, packaging and identification of living modified organisms (as GMOs are legally termed) across boundaries. Because the protocol seems likely to be ratified soon, we are very close to requiring internationally standardized documentation and instruction for transboundary movements of these organisms.

Global scientific communities need to pay careful attention to future practices, rules and standards in international movements of these organisms, even if they are being used only for research. We must, of course, have a proper method for considering potential risks to the environment and human health in handling, transport, packaging and identification. Also, it is vital to avoid confusion, unfounded fear and consequent negative perceptions of biotechnology associated with genetic engineering.

There is particular concern among Japanese scientists and industry about the present status of legislation, proposed by the government, on research and development (R&D) and industrial uses of living modified organisms, in the light of the biosafety protocol. Although commercial imports are unlikely to suffer from any new laws passed to meet consumer demands and concern, it may be impossible to regulate R&D and industrial applications, because of confusion about administrative responsibility and inadequate numbers of inspectors. Establishing comprehensive and workable rules under the protocol is urgent to avoid delays to R&D and commercial activities, which would set Japan back further in relation to North America and Europe.

So far, there has been no commercial cultivation of GM crops in Japan, although government-approved GM cultivars could be planted and sold to consumers without major regulations. Public concern in Japan discourages interest in commercial crop cultivation R&D. Work that has been carried out by academic research institutions and the private sector, and

past investment in genetic engineering, may be wasted because of these concerns and the impending regulations.

Japan has less experience in field evaluation of GM crops than countries in the European Union; there have been four times more evaluations in Italy and eight times more in France, where commercialization is unpopular but research is tolerated ([www.ois.oecd.org/biotrack.nsf](http://www.ois.oecd.org/biotrack.nsf)). In commercial cultivation of GM crops, Japan lags behind other Asian countries. India and China have far more experience at the local farmers' level, with thousands of hectares of GM crops and much accumulated knowledge and experience

for future development of crop technology and associated risk management.

Although there is high-quality basic research on plant molecular biology and genetic engineering in Japan, these valuable results have remained unexploited because of regulations and lack of support systems for implementing the environmental release of GM plants or a biosafety assessment. Japan needs regulatory agencies that have clear responsibilities, and clear, workable risk-management schemes in R&D institutions.

**Kazuo N. Watanabe**

Gene Research Centre, University of Tsukuba,  
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan

## Reviewing should be shown in publication list

**Sir**—M. H. Dominiczak's comments on peer review in Correspondence (*Nature* **421**, 111; 2003) are a useful reminder that this mainly secret activity is essential for the quality of scientific publications to be kept as high as possible. High-quality reviewing is time-consuming and in some ways is comparable to co-authorship, supervision and teaching, often giving rise to highly influential scientific debate. Yet it is almost without tangible reward, other than the private satisfaction of a job well done or the thanks of the author to an anonymous referee in the small print.

At present, very few scientific journals pay their reviewers. More substantial financial rewards are unlikely to improve the quality of reviews, and are too costly for most journals, as discussed in your News Feature "Publish, and be damned..." (*Nature* **419**, 772–776; 2002). For many years, as your feature noted, some journals have published lists of their referees, and the American Geophysical Union runs a scheme to honour their best reviewers. Yet although these initiatives are welcome, they do not provide lasting motivation.

Scientists often briefly mention their reviewing activity in their CVs, listing the journals and frequency of reviews. We propose that these qualifications should be made a standard part of the CV as well as of research units' annual reports. This offers relevant information about scientific status, acquaintance with the literature, and willingness to offer free advice, which in itself is of considerable relevance to grant applications and in appointments. The quality of the journals and the frequency of reviews would be a measure

of distinction, as editors use good reviewers most often. Our proposal could gradually improve the quality and status of reviewing—assuming that journals would be willing to verify reviewers' figures—and would provide a more enduring incentive for people to participate in this otherwise often frustrating duty.

**Torben Clausen, Ole Bækgaard Nielsen**

Department of Physiology, University of Aarhus,  
DK-8000 Aarhus C, Denmark

## Flying into history

**Sir**—The caption of your Commentary picture (*Nature* **421**, 16; 2003) suggests that Orville Wright made the first powered flight in 1903. This is not true, although Heilbron's and Bynum's text, on the previous page, was correct. The Wright brothers—influenced by the non-powered gliders of Otto Lilienthal, who flew more than 300 m in 1894—were the first to achieve the important conjunction of four criteria with their 260-m flight: it was manned, powered, heavier-than-air and (to some degree) controlled.

Earlier pioneers set records by meeting some of these criteria. In 1890 Clément Ader made the first manned, powered, heavier-than-air flight, of 50 m, in his bat-winged monoplane. Henri Giffard's steam-powered airship covered 27 km on the first manned and powered flight, in 1852. Balloonists Jean-François Pilâtre de Rozier and François d'Arlandes were, in 1783, the first men to fly. And if they were as fashionably bewigged as the occasion demanded, their 9-km ride must also have been the first manned, powdered flight.

**Jürgen Schmidhuber**

Dalle Molle Institute of Artificial Intelligence,  
Galleria 2, 6928 Manno (Lugano), Switzerland



# Toxicology rethinks its central belief

Hormesis demands a reappraisal of the way risks are assessed.

Edward J Calabrese and  
Linda A Baldwin

How clean is clean? Billion-dollar arguments on this question are common in the United States, as agencies — such as the Environmental Protection Agency (EPA) — face the need to remediate sites of uncontrolled or abandoned hazardous waste, Ground Zero being the foremost example in many minds. Likewise, debates rage about 'safe' levels of compounds in the body, for example lead-associated cognitive deficits in children, which are claimed to occur at blood lead levels lower than previously thought. In addition, the US Congress is exploring whether low doses of organic mercury preservatives are contributing to an apparent marked increase in childhood autism.

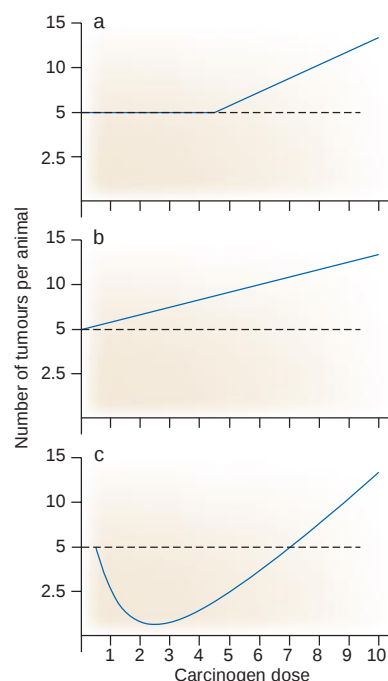
These, and numerous other examples, illustrate the central role that toxicology and the knowledge of the dose-response relationship play in a vast array of critical environmental, medical and public-health issues. As regulatory and public-health agencies base their decisions and policies on toxicological predictions, they are therefore of considerable importance to vast numbers of people as well as to national economies.

We believe the predictive models that all regulatory agencies use are based on a fallacy in the toxicological models used to predict and extrapolate dose responses from chemicals, pharmaceuticals and physical stressor agents. Here, we clarify the basis of this fallacy and advocate a more predictive model that will revolutionize public attitudes towards risk.

## Traditional models

The most fundamental concept used in toxicology to determine risk assessment and regulation is the dose-response relationship, for which two models have traditionally been used. The threshold model (Fig. 1a) is used in the assessment of risks for non-carcinogens, and the linear non-threshold (LNT) model (Fig. 1b) to extrapolate risks to very low doses of carcinogens. But we believe the most fundamental shape of the dose response is neither threshold nor linear, but U-shaped (Fig. 1c), and hence both current models, especially the linearity model, provide less reliable estimates of low-dose risk.

This U-shape is commonly called hormesis — where a modest stimulation of response occurs at low doses and an inhibition of response occurs at high ones<sup>1</sup>. The stimulation is often (but not always) observed following an initial inhibitory response, appearing to represent a modest



**Figure 1** Hypothetical curves depicting (a) threshold, (b) linear non-threshold, and (c) hormetic dose-response models using cancer (number of tumours per animal) as the endpoint. The reduction in number of tumours per animal at the lower doses (1–6) compared to the number of tumours per animal (5 tumours per animal) in the control indicates a reduced risk of cancer.

overcompensation to a disruption in homeostasis<sup>2</sup>. Depending on the endpoint that is measured, the hormetic dose response is either an inverted U — the endpoints being growth (such as the effects of low doses of various toxic metals, herbicides and radiation on plant growth) or survival (such as the effects of low doses of gamma rays on longevity in rodents) — or a J — the endpoint is the incidence of disease (for example, mutation, birth defects, cancer; Fig. 1c). Yet virtually all the leading toxicology textbooks consider only the traditional threshold and linear models.

The toxicological community made an error of historic proportions in its formative years (the 1930–40s) in buying into the threshold model<sup>3</sup>. Once accepted, this model became dogma, providing the basis for subsequent progress and confusion — despite toxicologists, radiation biologists, pharmacologists and others regularly pointing out unmistakable exceptions to the so-called threshold rule, such as the effects of saccharin,

dioxins, cadmium, mercury, numerous insecticides/herbicides, and numerous pharmaceutical agents. These unexpected results were generally written off either as reproducible but 'paradoxical' phenomena with no apparent capacity for generalization, or as biologically irrelevant random variation.

The implications of this systematic error are immense, not least in toxicological risk assessment. The *a priori* criteria we developed to assess whether experiments displayed evidence of hormesis based on study design, magnitude of the stimulatory response, statistical significance of the stimulatory response and reproducibility of findings, revealed up to 5,000 examples of hormetic responses independent of chemical class/physical agent, biological model and endpoint measured. Low levels of agents such as cadmium, dioxin, saccharin, various polycyclic aromatic hydrocarbons, X-rays and various gamma-ray sources reduce tumours in some species. Low doses of X-rays enhance life span in male and female mice and guinea pigs; ethanol and acetaldehyde enhance longevity in fruit flies; multiple stressor agents extend longevity in nematodes; numerous toxic substances (for example, cadmium and lead) enhance growth in various plant species. Low or modest consumption of ethanol reduces total mortality in humans, while increasing it at higher levels of consumption. The hormesis concept is thus highly generalizable and far-reaching.

Yet the vast majority of toxicological experiments are not designed to evaluate the hormetic hypothesis, assessing doses that are too high for the hormetic domain. Of those experiments that do have adequate study designs, a substantial proportion demonstrates hormesis. Using a database with rigorous and clearly defined entry and evaluative criteria, the hormetic model strikingly outperforms the 'dominant' threshold model<sup>4</sup>. The hormetic model is not an exception to the rule — it is the rule.

## Overlooking hormesis

So how did the field of toxicology get its most fundamental tenet, the nature of the dose response, so wrong? One reason is that, as mentioned above, most toxicological experiments lack the capacity to assess possible hormetic dose responses. Yet even when they do have potentially adequate study designs, the hormetic response can still be missed because at the assumed toxicological threshold dose (called NOAEL, for no observed adverse effect level), there is often evidence

of a low degree of toxicity, even if the response is not significantly different from the control group. As the dose below the standard threshold becomes progressively more dilute, the response becomes more likely to exceed the control value (hormetic-like). This is why mammalian toxicological studies, which emphasize high-dose toxicological responses such as those used to assess possible carcinogens in the US National Toxicology Program (NTP), are often incapable of adequately assessing the hormetic phenomenon. We believe that this combination of circumstances contributed significantly to the toxicological community overlooking the hormetic model and putting full emphasis on the threshold model for non-carcinogens and the linear model for carcinogens.

The mechanism by which hormesis occurs has also hindered its general acceptance. Toxicological researchers have rarely focused on why there are transitions (for example, stimulation followed by inhibition) in dose responses. Molecular pharmacologists, on the other hand, have focused on how such switching mechanisms work and how they affect the nature of the dose response, including hormetic-like biphasic dose-response relationships. There are more than 30 pharmacological receptor systems in the published literature that affect hormetic-like dose responses where the mechanisms that account for such responses have been clarified to at least receptor level<sup>5</sup>. These findings reveal that there is no single hormetic mechanism, but suggest a general strategy for resource conservation across biological systems.

Seven years ago, hormesis would not find its way into even informal conversation among toxicologists. Now, we not only know that it exists but accept its dominance over other models. The implications are enormous: they affect how toxicologists select biological models, choose endpoints to measure, design studies, assess risk and even pose the questions and the hypotheses they test. The dose response affects nearly all aspects of toxicological, pharmacological, epidemiological and clinical evaluation.

## Implications of hormesis

What are the implications of the hormetic perspective? Most notably, it challenges the belief and use of low-dose linearity in estimating cancer risks, and emphasizes that there are thresholds for carcinogens. The economic implications of this conclusion are substantial. The EPA has been struggling to harmonize how it assesses risks from non-carcinogens and carcinogens, having mistakenly assumed for a long time that non-carcinogens act via a threshold model whereas carcinogens act via a linear model at low doses. As both types of biological response follow the hormetic paradigm and display similar quantitative features of the

# Hormetic responses have equal, if not greater importance, for the biomedical and clinical sciences

dose response, the EPA could use the hormetic model as default to assess risk in both non-carcinogens and carcinogens<sup>6</sup>.

The hormetic perspective also turns upside down the strategies and tactics used for risk communication of toxic substances for the public<sup>7</sup>. For the past 30 years, regulatory and/or public-health agencies in many countries have 'educated' — and in the process frightened — the public to expect that there may be no safe exposure level to many toxic agents, especially carcinogens such as radiation and dioxins. If the hormetic perspective were accepted, the risk-assessment message would have to change completely. Changing a dominant risk-communication paradigm is not as simple as flicking on a light switch. It changes beliefs, attitudes, and assumptions, not unlike changing from a Soviet-style society to a western one. It would certainly be resisted by many regulatory and public-health agencies as an industrial-influenced, self-serving scheme that could lead to less costly, less protective clean-up standards, reminiscent of attempts by early opponents of hormesis to link it with homeopathy.

Hormetic responses have equal, if not greater, importance for the biomedical and clinical sciences. Many antibiotics, antiviral and anti-tumour agents, and numerous other medicines display hormetic-like biphasic dose responses: one dose may be effective clinically but another may be harmful. Some anti-tumour agents (for example, suramin) that inhibit cell proliferation at high doses, where they may be clinically effective, become like a partial agonist at lower doses, where they enhance cell proliferation. This is also true for some antibacterials (erythromycin and streptomycin, for example) and antiviral agents (such as gliotoxin analogues, colanolides, adefovir and Rhamnan sulphate). In these cases, the drug may be harmful to the patient at lower than therapeutic doses and requires careful clinical supervision. Some Alzheimer's treatments, such as the second/third-generation anticholinesterase agents, often enhance cognitive function at low doses but decrease it at higher doses. Thus, the hormetic biphasic dose response provides not only new opportunities for clinical improvements but also risks that have to be addressed.

Exercise is now being seen as a similar phenomenon, in that there may be an optimized degree of exercise that confers a wide range of benefits, whereas at higher levels (dose), the net result would be adverse. Immunology is likewise replete with examples of both chemical- and radiation-induced hormetic-like biphasic dose responses for a broad spectrum of endpoints and biological models. More than 150 endogenous agonists, drugs and pollutants induce hormetic effects in humans and other animals, affecting antibody production, cell migration, phagocytosis of microbes, destruction of tumour cells and other end-points. A better understanding of such phenomenon would have important implications for future research and biomedical development.

## Paradigm shift

At a time when the human genome has dominated many aspects of the scientific literature, it is generally unrecognized that the dose response of most, if not all, peptides conform to the hormetic model. Recognition of hormetic-like biphasic dose responses is important for elucidating the bioregulatory actions of various peptides and their biomedical implications.

Yet hormesis is not easy to study, as it requires the use of more doses (especially in the low-dose zone), often including a temporal component (measurement at various times within an experiment) and using more subjects to enhance statistical power, and needs replication. These extra features often steer researchers to less resource-intensive and more readily definable phenomena.

The hormetic dose response represents a paradigm shift in the concept of the dose response throughout biological science. It is widespread and outperforms other dose-response models. A general recognition of the hormetic perspective is likely to yield a vastly improved evolutionary basis of adaptive responses, scientific foundations of risk assessment and clinical medicine, as well as a more biologically plausible framework for understanding regulatory strategies at the level of the cell and the organism. ■

Edward J. Calabrese and Linda A. Baldwin are at the Department of Environmental Health Sciences, University of Massachusetts, Amherst, Massachusetts 01003, USA.

1. Luckey, T. D. *Radiation Hormesis* (CRC Press, Boca Raton, 1991).
2. Stebbing, A. R. D. *Mutat. Res.* **403**, 249–258 (1998).
3. Calabrese, E. J. & Baldwin, L. A. *Hum. Exper. Toxicol.* **19**, 2–31 (2000).
4. Calabrese, E. J. & Baldwin, L. A. *Toxicol. Sci.* **71**, 246–250 (2003).
5. Calabrese, E. J. & Baldwin, L. A. (eds) *Crit. Rev. Toxicol.* **31**, 349–669 (2001).
6. Callahan, B. G., Gaylor, D. & Stanek, E. J. (eds) *Hum. Ecol. Risk Assess.* **7**, 779–942 (2001).
7. Renn, O. *Hum. Exper. Toxicol.* **17**, 431–438 (1998).

## Acknowledgment

Our work is sponsored by the Air Force Office of Scientific Research. Our views do not necessarily represent those of the Air Force Office of Scientific Research or the US Government.

# Testing time for Russian lessons

Mathematicians say Russia's education reforms are a formula for disaster.

*Obrazovaniye, Kotoroye My  
Mozhem Poteryat (The Education  
That We May Lose)*

editor-in-chief V. A. Sadovnichii  
*Moscow State University Press/Institute for  
Computer Research: 2002. 288 pp.  
300 roubles*

Valery N. Soyfer

The Russian Ministry of Education plans to lengthen high-school education from 11 to 12 years and to replace university entrance exams with a unified examination system. The students' grades in these final high-school examinations will determine the level of funding that their universities will subsequently receive from the state. In September last year, the balance of subjects taught in high schools changed, with less time spent on mathematics and natural sciences, and more on the social sciences, information technology, local geography and ecology, physical education and, more worryingly, patriotic education and military training.

Most education specialists are opposed to these changes. A group of leading mathematicians, headed by Victor A. Sadovnichii, president of Lomonosov Moscow State University, has openly taken a stance against the reforms by publishing this book.

In a recent speech, the text of which opens this book, Russian President Vladimir Putin stated in general terms the importance of education and the need for reform. In their introductory article, Sadovnichii and Zhores Alferov, who won the 2000 Nobel Prize in Physics, consider in detail how such reform can be brought to education. They object to "attempts to impose quasi-market mechanisms" on the education system, and to the commercialization of high-school and university education.

Vladimir Arnold calls the proposed reforms a "preparation for a new Cultural Revolution" reminiscent of the crushing of Chinese culture and education by the communist government of Mao Tse-tung. He is concerned that the damage to Russia's higher education and culture would be long term and difficult to correct.

Of course, these reforms are not the first to have been attempted in the past hundred years. During the 75 years of communist rule there were several attempts to change the high level of education that existed in Russia under the tsars, but nevertheless, until recently, the traditions of one of the best secondary-education systems in the world were maintained. Sadovnichii, in a central article, points out that education traditionally "distinguished Russia in Tsarist times,



A word in your ear: Victor Sadovnichii (right) urges President Vladimir Putin to rethink on education.

the Soviet period, and today... The virtues of the Russian higher school, which the entire world spoke about with real respect, have always depended first of all on fundamental science, and on the scientific schools."

Examining the history of educational reforms in Russia in detail, Lev D. Kudryavtsev concludes that the innovations proposed — particularly the reduction in the level of teaching in mathematics and the natural sciences — may have disastrous effects on the country's future. He also expresses concern about the linking of university funding to the grades obtained by students in uniform state exams, given "the level of corruption that exists in our modern society".

The book is given emotional impetus by the transcript of a television interview given by Nobel laureate Aleksandr Solzhenitsyn, who worked for many years as a high-school teacher. He was concerned that the lack of state funding has led to the delapidation of many schools, particularly in rural areas. This, coupled with the high mortality rate in these areas, may mean that by 2010 "the number of school-aged children will have decreased by 30 percent... This means that the smaller rural schools will close. In the expanses of Russia the hearths of education will cease to glow."

Another former high-school teacher, Igor F. Sharygin, argues that mathematics should become a cornerstone of Russian education. He says that people who are mathematically literate and who understand what proof means cannot easily be manipulated: "Math-

ematics and state authority are two incompatible things, but rational sovereigns often run to mathematicians to resolve the most varied problems in difficult moments."

To substantiate their arguments, the editors have included a full translation of a report from the US National Commission on Mathematics and Science Teaching (the John Glenn Commission), along with the text of a programme of educational reforms proposed by US President George W. Bush. According to these documents, which take up nearly 100 pages, the US leadership is determined to counter the decline in the standard of mathematics and science teaching in US secondary schools. The editors of this book point out that while the United States is trying to improve its standard of teaching, the Russian government is taking measures in the opposite direction.

The publication of this book strikes me as an unusual step. Sadovnichii, the book's editor-in-chief, is one of Russia's best academic administrators, and the fact that he has taken to open and dispassionate conflict with the ministry of education is symbolic of a genuine civil society developing in Russia. It also provides evidence that the interests of society are being taken up by the most important leaders, rather than being concealed within the offices of the bosses, as happened in the Soviet period. Sadovnichii and his co-authors are to be applauded for their civil and open position. ■

Valery N. Soyfer is at George Mason University, Fairfax, Virginia 22030, USA.

A. SAZOVY/AP



## Looking up to the stars

**On the Shoulders of Giants: The Great Works of Physics and Astronomy**

edited, with commentary and introduction, by Stephen Hawking  
Running Press: 2002. 1,266 pp. £20, \$29.95  
Owen Gingerich

I am in two minds about this hefty tome. Should I be embarrassed for Stephen Hawking because an enterprising publisher has inveigled him into putting his name to a collection of superseded texts? Or should I be outraged that an eminent scientist, but one with no track record in the history of science, has the arrogance to endorse historical introductions for five classics of science?

*On the Shoulders of Giants* comprises English versions of texts from five distinguished cosmologists — Copernicus, Galileo, Kepler, Newton and Einstein — each preceded by a 6–8-page commentary on the life and work of the author. Included in these texts

are Copernicus' paean to the heliocentric system, Galileo's formulation of projectile motion, Kepler's harmonic law, Newton's law of gravitation, and Einstein's equation  $E = mc^2$ . However, these key passages are not highlighted: they are buried in long, technical treatises with no clue as to where to locate them. Neither will you find Galileo's arguments for a heliocentric system, nor Kepler's law of areas or ellipses, because the appropriate texts are not included.

Each translation has apparently been chosen for its easy availability, and each now has a more nuanced version currently in print elsewhere. Take Newton's *Principia*, for example. For years the standard English version was Florian Cajori's modernization of Andrew Motte's 1729 translation. With the appearance of I. B. Cohen's magisterial translation in 1999, the University of California Press withdrew Cajori's version. *On the Shoulders of Giants* reverts back to Motte's dated translation.

The choice of Galileo's text, here titled 'Dialogues Concerning Two Sciences', is particularly curious. Modern scholars call the book *Discourses* or *Two New Sciences* to distinguish it from his *Dialogue on the Two*

*Great World Systems*, which was Galileo's classic defence of heliocentric cosmology that got him into trouble with the Inquisition. The dust-jacket mentions that one of the texts included was considered so "dangerous" that its author was accused of heresy, as if the *Dialogue on the Two Great World Systems* were found here. This also seems to be what Hawking expected, because in his introduction he writes: "In *Dialogues Concerning Two Sciences*, Galileo's characters, Salviati and Sagredo, put forward persuasive arguments in support of Copernicus." So they do in the *Dialogue on the Two Great World Systems*, but this work is not included here.

Kepler's *Harmonices mundi libri V* is an especially inept selection. This fifth book does include a splendid nugget — Kepler's mathematical relationship between the cube of the planet's distance from the Sun and the square of its period — but it is an idiosyncratic work, here presented entirely out of context. And its title is erroneously translated as 'Harmonies of the World'. *Harmonices* looks like a Latin plural, but Kepler used a Greek genitive singular ending, because, as a great unifier, he believed in a singular 'Harmony of the World'.

Kepler's *Astronomia nova* ranks beside Copernicus' *De revolutionibus* and Newton's *Principia* in the triumvirate of technical landmarks in the scientific revolution, but it's not here. And it would have been a redeeming achievement to have included William Donahue's scarce translation of Kepler's *New Astronomy* in this collection instead of part of *Harmony of the World*.

The commentaries on each author contain congenial anecdotes, some accurate historical data, and a dose of modern mythology. Unfortunately, the opening vignette is littered with errors. Hawking stumbles in the very first line, calling Copernicus a priest, which he most surely was not, although he was a canon and legal officer at Frauenburg cathedral. Hawking is in good company here, because Galileo also described Copernicus as a priest, but Galileo had propagandistic reasons that Hawking lacks.

According to Hawking, Aristotle argued that the Earth was round because hulls of ships sailing out to sea disappeared over the horizon before the sails. The argument was made by Ptolemy half a millennium later, but not by Aristotle. Hawking's introduction also says that Western Christendom placed Hell beyond the stars; that Copernicus became a professor of astronomy at Bologna; that he completed *De revolutionibus* in 1530; that Rheticus relinquished a chair in mathematics at Wittenberg to study under Copernicus; that Copernicus used equants to account for the motion of the Earth; that Osiander placed the word 'hypothesis' on the title page of Copernicus' book; and that the world had scarcely become known to be round when Copernicus wrote. None of this is true. No



## A brush with nature

While out walking in the Florida Everglades, wildlife painter Robert Bateman saw a flash of pink. Closer inspection revealed it to be a group of roseate spoonbills, swinging their heads from side to side as they fed on crustaceans.

The resulting painting (above) is one of 200 new works in his book *Birds* (Pantheon, \$40). The paintings are accompanied by short accounts of Bateman's adventures as he sought out his subjects in their natural environments.

bibliography is provided, so it is difficult to ascertain the source of this disaster. Fortunately the other introductions are not as bad: the Kepler introduction benefits from using the *Dictionary of Scientific Biography* as its unacknowledged source, for example.

The most interesting part of this book is the general introduction; this is quintessential and thoughtful Hawking, clearly carrying his own stamp. He writes about the anthropic principle: "If the ultimate theory made a unique prediction for the state of the universe and its contents, it would be a remarkable coincidence that this state was in the small subset that allows life." It is almost worth the price of the book to get this quotation. ■

Owen Gingerich is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

## A trick of the moonlight

**The Mystery of the Moon Illusion**

by Helen Ross & Cornelis Plug  
Oxford University Press: 2002. 277 pp.  
£29.95, £45

Maurice Hershenson

When the Moon is near the horizon it appears to be larger and closer than when it is high in the sky. Yet light reflected from the Moon to the eye of an observer on Earth provides the same stimulus, regardless of the elevation of the Moon. This phenomenon is known as the 'Moon illusion'. A similar illusion is observed for the setting Sun and for celestial distances between star points at different elevations.

The Moon illusion is arguably the oldest unsolved problem in modern science. It is mentioned in cuneiform script on clay tablets from the royal library of Nineveh and Babylon, dating from before the sixth century BC, and in a collection of Chinese legends ascribed to the Taoist philosopher Lieh-tzu, dating from the fifth century BC.

Until the past hundred years or so, the illusion was thought to be a consequence of physical processes. For example, Aristotle (in the third century BC) and Ptolemy (in the second century AD) incorrectly attributed the illusion to the magnifying properties of the atmosphere, and Alhazen (Ibn al-Haytham) related the illusion to the flattened appearance of the dome of the sky.

Throughout its long history, the illusion has been considered by many of the leading scientists and mathematicians of their day: da Vinci, Kepler, Descartes, Huyghens, Euler and Reimann, to name but a few. In *The Mystery of the Moon Illusion*, Helen Ross and Cornelis Plug carefully guide us through the history of explanations of the illusion,



Moonrise over the savannah: but just why does the Moon appear larger when it is near the horizon?

treating in turn each aspect of the stimulus situation and its context.

Over the past century it has become clear that the Moon illusion is psychological, a consequence of the processes that underlie visual perception. Unfortunately, the authors do not provide as clear a picture of the treatment of modern explanations, perhaps because there is little agreement among current workers. For example, many modern explanations invoke the size-distance invariance hypothesis (SDIH), by which the stimulus (measured as the angular subtense of the light reaching the eye, or the visual angle) determines how large an object looks (its perceived size) and how far it appears to be from an observer (its perceived distance). The hypothesis is embodied in a psychophysical equation relating the physical stimulus (input of the visual angle) to the ratio of perceived size to perceived distance. Some theorists treat the SDIH as a fundamental law of visual perception. Indeed, it is this relationship that makes the Moon illusion such a puzzle, because it requires that the Moon looks farther away if it appears larger, and smaller if it appears closer. This is also why the illusion has sometimes been said to be paradoxical.

Modern explanations of the Moon illusion propose major changes in the form of the SDIH. For example, Walter C. Gogel and Davis L. Mertz retain the traditional form but allow perceived distance to be referenced simultaneously by different behavioural responses. Thus, the Moon may actually appear more distant to people who say: "The Moon looks close." Lloyd Kaufman and Irvin Rock's explanation is similar except that they do not include perceived distance in their formulation. Instead, they suggest that 'registered distance' (information about distance) enters the SDIH to determine perceived size. This approach alters the

relationship because distance is now input information, rather than distance perception being output. Verbal statements about the perceived distance of the Moon are described as inferences based on perceived size, rather than descriptions of experience: "The Moon looks big so it must be close."

Don McCready altered the SDIH even more by substituting a perceptual outcome (the perceived visual angle) for the stimulus input (the visual angle). This proposed relationship is not just another form of the traditional psychophysical formulation, but a purely psychological relationship.

Finally, in my own explanation, the SDIH becomes a special case of a more general mechanism — a kinetic SDIH — that governs the perception of rigid object motions in depth. This mechanism automatically transforms changing stimulus inputs into objects that appear to be rigid but moving in three-dimensional space. When the stimulus does not change, as in the Moon illusion, the perceived distance of the object is determined by contextual stimulus information, such as the ground and horizon. The perceived size of the object is determined by distance information processed by the mechanism that produces rigid object motions. In this case, an object that appears close in distance, such as the Moon at the horizon, is made to appear large in size.

Despite the vast history of research into the Moon illusion, so ably reviewed by Ross and Plug, there is still no single explanation upon which even a handful of workers agree. I am left to conclude, as I did in 1989, that more research is needed, but not on the Moon illusion. An explanation of that heavenly paradox will be obvious once our understanding of visual space perception is clear. ■

Maurice Hershenson is in the Department of Psychology, Brandeis University, Waltham, Massachusetts 02254, USA.

# A thirst for knowledge

*Summarize yourself in the form of a title of a paper in Nature.*

Evidence for the emergence of order from chaos.

*What was your first experiment as a child?*

A basic newtonian physics experiment on mass and gravity in which I pulled a wall down on top of myself, sustaining a leg wound that needed more than 50 stitches and a painful eight-month convalescence. I concluded that medicine is not an experimental science.

*Who has been the most important mentor in your career?*

Eraldo Antonini and Alessandro Finazzi-Agro' from Rome and — much later, upon appreciation of their professional achievements — my mother and father.

*What makes a good scientific mentor?*

Lots of discussion, and most importantly patience and appreciation when the students don't listen to advice. My common experience is that the best ideas originate from discussion, and are generated in the pub, providing a forum, an atmosphere within the group where creativity can flourish. But the data require extreme rigor.

*What gives you the most job satisfaction now? What are your major frustrations?*

My greatest satisfaction is designing new experiments. My greatest frustration is trying to write their results into original papers.

*What single scientific paper or talk changed your career path?*

Conversations with Jeffries Wyman, attempting to explain the relationship between nature, beauty, symmetry and the allosteric effects of haemoglobin.

*What's your favourite conference destination, and why?*

Keystone, Colorado, because of the mountains, the skiing, the people, the science and the time to discuss it.

*You have the audience in your hands, but some smart-alec asks you the killer question that you have no idea how to answer. What's your favourite response?*

"Whereas a first analysis seems to accord with your suggestion, a closer investigation reveals unexpected levels of complexity."

*What was the worst/most memorable comment you ever received from a journal?*

Receiving a letter of acceptance, and later the galley proofs, of a paper for somebody else; I was strongly tempted to correct them,

adding my name as co-author. I also enjoyed being asked to act as referee for my own paper. (I declined.)

*What book is currently on your bedside table?*

Aristophanes' *The Wasps*. It is incredible how modern it is, even after 24 centuries.

*Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?*

True to my Italian roots I'd go for a large dinner with plenty of room for discussion: there'd be Socrates to discuss whether knowledge has advanced at all since his time; Galileo to ask what he thinks of patents; Julius Caesar to ask what he thinks of the European Union; Pericles to hold forth on the modern balance between technology, politics, the military and lawyers; Vivaldi to comment on Eminem's harmonic structure; Michelangelo to discuss the 'cut-and-paste' nature of modern creativity; Dante to update his *Divina Commedia* with most recent important people; Wittgenstein to dismantle any concept brought forth by these encounters; Sappho to report the event; the *Monty Python's Flying Circus* team to film it — but, most important, Bacchus to organize dinner.

*You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?*

I would introduce myself as 'my arch-rival', criticize my methods and generally discredit myself with lies and vicious slander, and see how they reacted. Then of course I'd introduce the real 'me', do a proper criticism of my work and invite them all for a good session in the pub.

*What do you most dislike about having research published?*

Having to read it a few years later.

*The Internet is the bane of scientists' lives because...*

...we can't think quickly enough! The net is a very powerful instrument, allowing more and faster communication, especially at a distance. Science pretends to reach truth (like art) through objectivity (unlike art): the findings of the individual must be ratified by the consensus of the many, and hence communication in the form of meetings, conferences, published papers and so on is indispensable. Science is a recent manifestation of a general class of discourse known as 'representation'. Older examples of the genre are labelled either sacred or profane, and are performed in churches and theatres, respectively. No surprise then that some scientists act like priests,



## Gerry Melino

Gerry Melino works on apoptosis — programmed cell death — concentrating on the role of p73 in cancer and in keratinocytes. He is at the University of Rome 'Tor Vergata' but is relocating to the Medical Research Council's Toxicology Unit in Leicester, UK.

and others like actors! In all cases a plurality is required to expand the discussion; here the net is fantastic. But in our line of work some time is needed for concepts to sink in and for the full ramifications of findings to be contemplated. No amount of flashy technology beats what each of us carries between our ears.

*What do you do to relax? (Please bear in mind that Nature is a family journal.)*

What do you mean 'family journal'? I'm as pure as the driven snow!

*Harry Potter or Lord of the Rings?*

Well, Harry Potter has the same dreams, fantasies, experiments and problems of communication as many scientists...

*What would you have become, if not a scientist?*

An artist. Science and art are fundamentally imaginative. The popperian idea of science is based on conjecture and refutation. But the conjecture in science, as in art, necessarily comes first.

*What's just around the corner?*

The beginning of molecular research on memory and learning.

*Name one extravagance you can now get away with because of your eminence.*

I could never match Luc Montagnier, whose house was broken into and a number of things stolen. But when the robbers saw his name on the credit cards, they recognized it as the name of the man who discovered HIV, and returned everything. His house was never broken into again. Now *that's* eminence. ■



# The current climate

Stefan Rahmstorf

How are ocean currents caused, and how do they affect climate? These questions were hotly debated in the nineteenth century. Some argued that water is simply pushed along by the wind; others postulated 'convection currents' caused either by heating and cooling or by evaporation and precipitation. Even today, the driving forces and climatic effects of ocean currents are still not completely understood.

In 1908, Johan Sandström laid the foundations of our modern understanding of ocean currents with a series of classic experiments carried out at Bornö oceanographic station in Sweden. He filled a small tank with water layers of different densities from an adjacent fjord and then blew air over the surface and/or heated and cooled the fluid at different levels. He thus elucidated the properties of 'wind-driven' and 'thermal' circulation. The latter term was amended by the 1920s to 'thermohaline circulation', because water density in the ocean is determined by both temperature and salinity.

Sandström found that thermal forcing can give rise to a steady circulation only if heating occurs at a greater depth than cooling — a fact that is familiar to oceanography students as 'Sandström's theorem'. But 'thermohaline forcing' — that is, fluxes of heat and freshwater — occurs only at the ocean's surface, except for a small contribution from geothermal heating.

So what is the deep heat source that drives the ocean's observed thermohaline circulation? Sandström recognized that it is the

downward penetration of heat at low latitudes, due to turbulent mixing, that provides this thermal engine. His speculation that salinity-driven convection causes the low-latitude mixing was wrong, however — turbulent mixing is instead powered by winds and tides. Thermohaline circulation is thus caused by the joint effect of thermohaline forcing and turbulent mixing — it can be defined as currents driven by fluxes of heat and freshwater across the sea surface and subsequent interior mixing of heat and salt. Although winds and tides are important in creating turbulence, this driving mechanism is clearly distinct from wind-driven circulation: thermohaline circulation requires thermohaline surface forcing, whereas wind-driven circulation does not.

To complicate matters, the ocean's density distribution, which determines pressure gradients and thus circulation, is itself affected by currents and mixing of any kind. Thermohaline and wind-driven currents cannot therefore be separated by oceanographic measurements. There are thus two distinct forcing mechanisms, but not two separate circulations. Change the wind stress, and the thermohaline circulation will change; alter thermohaline forcing, and the wind-driven currents will also change. It is because of thermohaline forcing that wind-driven currents are relegated to the upper ocean — in unstratified water they would extend to the bottom.

But although thermohaline circulation is not measurable, the concept is still a useful one, and modern computer models of oceans can be used to carry out experiments (not unlike those of Sandström with real water) to study the properties of these currents. In these models, different surface forcing fields can be prescribed, and by designating the surface wind-stress as zero, a purely thermohaline circulation can be computed. The required turbulent mixing can be varied independently from the surface wind stress, as they appear in different terms of the hydrodynamic equations. The resulting zonally integrated circulation is essentially similar to the circulation obtained with wind-stress forcing, but lacks the wind-driven cells known as Ekman cells (which consist of surface water that is pushed along by the wind and returns within the upper few hundred metres of the ocean).

In this sense, zonally integrated streamlines can be interpreted as a superposition of wind-driven and thermohaline components, much as Sandström interpreted his tank experiments. On the other hand, when wind stress remains constant, the vertically integrated circulation looks similar with or without thermohaline forcing, with the exception of the Antarctic Circumpolar Current.

Wind-driven and thermohaline circula-

## Thermohaline circulation

*Heat and freshwater fluxes at the ocean's surface play a key role in forming ocean currents, which in turn have a major effect on climate.*

tions can thus be disentangled to some extent by models, which helps us understand what aspects of the circulation are linked to what surface boundary conditions. This is useful in analysing the effect of a change in forcing, such as a freshwater influx, on currents — a typical problem in palaeoclimate studies, in which sediment data suggest that freshwater influx has caused major changes in currents in the past. It is also highly relevant when considering the ocean's response to global warming, because evaporation, precipitation and runoff are expected to increase in a warmer world.

How strongly might changes in thermohaline circulation affect climate? To what extent do Europe's mild winters depend on the transport of heat by the Gulf Stream and North Atlantic Current? Simulations in which the ocean's heat transport is switched off consistently show a large winter cooling over the northern Atlantic and adjacent land areas, reaching several degrees in inland Europe, up to 10 °C over Greenland and even exceeding 20 °C over the Nordic seas. This heat transport warms the climate on both sides of the Atlantic, and is therefore not the main reason that Europe is warmer than Newfoundland — this phenomenon is mainly due to the prevailing winds in the two regions. But ocean currents do make the northern Atlantic much warmer than at comparable latitudes in the northern Pacific. Changes in these currents are our best explanation for the abrupt and marked climate swings that occurred over the north Atlantic many times during the most recent glacial period, as shown by Greenland's ice cores and by deep-sea sediments. Circulation changes might again be triggered by global warming.

Despite its known limitations, the concept of a thermohaline ocean circulation remains well defined and useful. Understanding its past and future behaviour is crucial to our understanding of climate change.

Stefan Rahmstorf is at the Potsdam Institute for Climate Impact Research, Box 601203, 14412 Potsdam, Germany.

### FURTHER READING

Rahmstorf, S. *Nature* **419**, 207–214 (2002).  
Seager, R. *et al.* *Q. J. R. Meteorol. Soc.* **128**, 2563 (2002).  
Vellinga, M. & Wood, R. A. *Clim. Change* **54**, 251–267 (2002).  
Wunsch, C. *Nature* **405**, 743–744 (2000).

S. ANDREUETTE, MULLER-KARGER/UNIV. S. FLORIDA/INST. MAR. REMOTE SENSING

**Wave review:** ocean currents sculpted these sand and seaweed formations, dozens of miles across.

# A magnificent machine

Richard B. Vallee and Peter Höök

Electron-microscope studies of the motor protein dynein reveal fascinating details of the movements of its various structural regions. The protein displays a degree of gymnastic ability that is rarely seen.

Only a handful of proteins have the remarkable ability to use the chemical energy stored in ATP molecules to produce force and carry out work. Dynein is the largest of these so-called motor proteins, and, because of its complexity, the mechanism by which it produces force has been a mystery. But on page 715 of this issue, Burgess and colleagues<sup>1</sup> provide the most detailed picture to date of how this motor works. Their computer-averaged electron-microscopic images show the protein in two stages of its force-production cycle, and the proposed mechanism of action is unique.

Dynein produces force against the surface of microtubules — long, hollow polymers that occur in nearly all cells. One type of dynein powers cilia and flagella: these structures have a bundle of microtubules at their core, and oscillate at high frequency by virtue of the force generated by the dynein molecules arrayed along the microtubules. Another form of dynein, cytoplasmic dynein, binds to diverse subcellular structures, including membranous organelles, chromosomes, the inner surface of the cell membrane and at least some viruses. Through these associations, dynein controls many forms of cellular and intracellular movement.

To drive these processes, dynein must grab hold of a microtubule, hydrolyse ATP, and then convert the chemical energy released into mechanical energy that can be used to push on the microtubule and move along it (or move it along). Then the cycle, which occurs many times per second, begins again. This process relies on conformational changes within dynein — but the details of these changes, and how they generate force, are largely unknown.

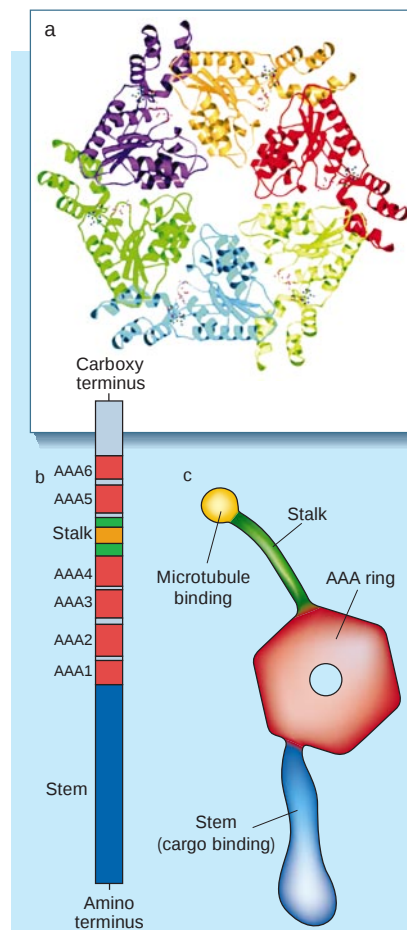
Comparisons with other proteins can be instructive when trying to understand a protein's structure and mode of action. However, comparing dynein with another class of microtubule motor protein, kinesin (Box 1, overleaf), has yielded few clues. Kinesin's motor domain — the part of the molecule responsible for ATP hydrolysis and force production — has a modest relative molecular mass of 35,000, and binds microtubules through amino acids on its globular surface<sup>2</sup>. In striking contrast, dynein's motor domain is ten times larger at a relative molecular mass of 350,000. And the protein is evolutionarily distinct from kinesin, instead, rather surprisingly, being related to the AAA

family of ATP-hydrolysing enzymes<sup>3,4</sup>. These enzymes participate in diverse functions, apparently unrelated to that of dynein, such as protein folding, unfolding and disassembly, and DNA replication. But many of them adopt a common, hexameric ring-like structure (Fig. 1a). (At least some of the activities of the AAA proteins may involve the production of force, but whether this is indeed the case, and how conformational changes within the rings might contribute, is incompletely understood<sup>5</sup>.)

The defining feature of the AAA proteins is a structural region called the AAA module. Dynein itself contains four evolutionarily well-conserved AAA modules, and two marginally conserved AAA modules, all within a single, long polypeptide subunit known as the heavy chain, which has a relative molecular mass of more than 500,000 (Fig. 1b). In this regard, dynein is quite different from most AAA-family members, which are composed of six small and identical subunits. Nonetheless, evidence obtained by averaging electron-microscopic images of individual dynein motor domains has revealed a ring-like organization<sup>6</sup> (Fig. 1c), consistent with the general AAA structural plan.

The dynein motor domain also has another curious yet striking feature, a fine stalk that links the AAA ring to microtubules. Predictions based on the amino-acid sequence of the stalk suggest that it consists of an unusual projecting  $\alpha$ -helical polypeptide strand, folded back on itself, with a small globular microtubule-binding site at its tip<sup>7</sup>. As dynein cycles through from microtubule binding to ATP hydrolysis to force production, information on the changes occurring in force generation and conformation must be transferred through the stalk — but how this is achieved is unknown. In addition, at the base of the dynein molecule there is a more substantial extension, the stem, which anchors dynein to one subcellular structure or another. Unlike the stalk, the stem was not expected to have an active role in force production.

This is where Burgess *et al.*<sup>1</sup> come in. They looked at a form of dynein that is found in the flagella of the alga *Chlamydomonas reinhardtii*. By analysing whole dynein molecules, the authors could take advantage of the stalk and stem, using them as landmarks. This allowed them to obtain average images of subsets of dynein molecules in several



**Figure 1 Understanding dynein.** a, Crystal structure of part of the AAA protein NSF. The defining feature of such proteins is a structural region called the AAA module. Here, a hexamer (diameter ~12 nm) of such modules is shown, with different colours distinguishing individual modules. (Image reproduced from ref. 10.) b, The sequential arrangement of structural regions, including six AAA modules, within the heavy-chain polypeptide subunit of dynein. c, Morphology of dynein. The AAA modules form a ring akin to that found in more conventional AAA proteins such as NSF. The hydrolysis of ATP induces changes within this ring that are passed along to both the stem (which anchors dynein to diverse subcellular structures) and the microtubule-binding stalk. This results in force production and movement. Burgess *et al.*<sup>1</sup> have described, in unprecedented detail, computer-averaged electron-microscopic images of dynein at two stages of this process.

## Box 1 A trio of molecular motors

The term 'motor protein' was introduced some 15 years ago to describe proteins that produce force. Although it is being expanded, the definition originally referred to three classes of protein: the myosins, kinesins and dyneins. These proteins are particularly important, as they now appear to be responsible for most forms of cellular movement.

All three types of protein use ATP as an energy source. Like most enzymes, they undergo conformational

changes during their catalytic cycle. What makes them unique and remarkable is their ability to couple these changes to net movement along fibres within cells. The myosins were initially identified for their role in muscle contractility, but they are now known to be responsible for cellular 'crawling' movements and other functions. The kinesins are the most recently identified class, and are needed for, among other tasks, the transport of

secretory and synaptic vesicles.

The dyneins were first discovered to underlie the movement of cilia and flagella, but are now also known to transport membranes, chromosomes, viruses, and various organelles and proteins. Over the years, our understanding of force generation by myosins and kinesins has extended to the atomic level. How the dyneins cause movement is one of the remaining frontiers in motor-protein biology. **R.B.V. & P.H.**

orientations, dramatically improving the level of detail discernible.

The images confirm the ring-like appearance of the motor domain, but show that it is surprisingly asymmetric. Although individual blobs can be seen that have the dimensions of AAA modules, there is no orientation of the molecule that allows all of these units to be counted together<sup>1</sup>. In addition, images of the two faces of the ring are clearly distinct. Together, these observations are consistent with evidence that the structure of the individual dynein AAA modules has diverged during evolution<sup>3</sup>.

In these images, the stalk emerges with unprecedented clarity; even the two strands of  $\alpha$ -helical structure at the stalk's base can be discerned. Unexpectedly, the stalk is bent. This was not seen in earlier studies using fast-frozen dynein molecules<sup>8</sup>, and could conceivably result from the molecule's exposure to the harsh conditions of negative-stain electron microscopy. Yet it is sufficiently reproducible to persist in the averaged images.

Of greatest interest are the changes seen in dynein in its different enzymatic states. Dynein hydrolyses ATP to produce ADP plus phosphate. It is thought to exist in a cocked state at this point, and to undergo its 'power stroke' on subsequent release of ADP<sup>9</sup> (which will allow a new ATP molecule to bind and the cycle to start again). Burgess *et al.* examined dynein bound with ADP plus vanadate — a dead-end complex representing the state just after ATP hydrolysis and just before the power stroke — and compared these images with those of the molecule after its power stroke (that is, in the absence of both ATP and ADP). The conformation is seen to change substantially, with the most dramatic feature being a shift in the angle between stalk and stem. If a dynein molecule were fixed at its base, this alteration would result in a 15-nm movement of the tip of the stalk.

In what part of the dynein molecule does

the primary change in conformation occur to generate this power stroke? Here the evidence is indirect, but very suggestive. In a few of the dynein molecules studied by Burgess *et al.*, the stem region is some 10 nm longer than in most other images. The functional state of these molecules is unknown; they may even be inactive. Nonetheless, the extra segment of protein — which the authors term the linker — must be present somewhere within the other more numerous images. The authors argue that this is indeed the case, and that the linker is kinked and twisted behind the AAA ring in those images. This unconventional hypothesis is supported by the variable appearance of the 'hole' in the ring, which is clearer in one state of the enzymatic cycle, and occluded in the other. It is also more evident when seen from one face of the ring than from the other. Perhaps, the authors argue, power is generated through a change in the interaction of the linker with the AAA ring. In addition, the stalk becomes straighter after the power stroke. Whether

this change contributes to the power stroke is uncertain, but it may help to explain how conformational information is transmitted to and from the microtubule.

These findings<sup>1</sup> predict a level of gymnastic ability that is rare among proteins. During its presumed evolution from more conventional AAA proteins, dynein appears to have acquired two prominent projections — the stalk and stem — each of which behaves unusually. The stalk alone may be capable of conformational flexibility not generally found in comparable  $\alpha$ -helical structures. The proposed behaviour of the newly discovered linker domain is remarkable and will also bear further examination. Curiously, the extended version of dynein seen occasionally by Burgess *et al.* may be the more common view seen in earlier studies<sup>8</sup>. How these observations can be reconciled, and how changes in dynein conformation relate to movement along microtubules, are issues for the future. ■

Richard B. Vallee and Peter Höök are in the Departments of Pathology, and Anatomy and Cell Biology, and the Center for Neurobiology and Behavior, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA.

e-mail: rv2025@columbia.edu

1. Burgess, S. A., Walker, M. L., Sakakibara, H., Knight, P. J. & Oiwa, K. *Nature* **421**, 715–718 (2003).
2. Vale, R. D. & Fletterick, R. J. *Annu. Rev. Cell. Dev. Biol.* **13**, 745–777 (1997).
3. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. *Genome Res.* **9**, 27–43 (1999).
4. Mocz, G. & Gibbons, I. R. *Structure* **9**, 93–103 (2001).
5. Ogura, T. & Wilkinson, A. J. *Genes Cells* **6**, 575–597 (2001).
6. Samso, M., Radermacher, M., Frank, J. & Koonce, M. P. *J. Mol. Biol.* **276**, 927–937 (1998).
7. Gee, M. A., Heuser, J. E. & Vallee, R. B. *Nature* **390**, 636–639 (1997).
8. Goodenough, U. W. & Heuser, J. E. *J. Mol. Biol.* **180**, 1083–1118 (1984).
9. Holzbaur, E. L. & Johnson, K. A. *Biochemistry* **28**, 5577–5585 (1989).
10. Yu, R. C., Hanson, P. I., Jahn, R. & Brunger, A. T. *Nature Struct. Biol.* **5**, 803–811 (1998).

### Quantum gravity

## The quantum of area?

John Baez

What is the smallest unit of area? To find out, theorists have been wrestling with the notion of quantum black holes. Two independent analyses now seem to lead to the same answer.

**A**mong the many curious features of quantum mechanics, one of the first to be discovered was that atoms have a discrete spectrum of energy levels. Could the same be true for black holes? To know for sure, we would need to understand how quantum mechanics and general relativity fit together — one of the great unsolved problems in physics. But we can already make a guess at the answer, just as Bohr guessed a

lot about the spectrum of hydrogen before quantum mechanics was fully worked out. Now, with the publication of Olaf Dreyer's work<sup>1</sup> in *Physical Review Letters* — and much to the shock of the experts — it seems that two completely different ways of guessing have arrived at the same answer. Starting from a black hole with a spectrum of discrete energy levels and relating that to its surface area, both agree on the size of the quantum of



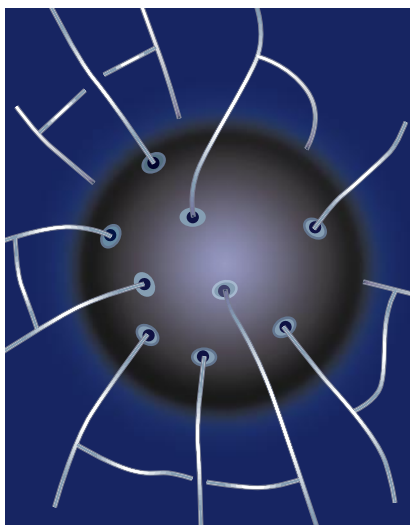
area, although no one yet understands why.

It all started in 1975, when Hawking<sup>2</sup> predicted that black holes aren't quite black. Taking quantum mechanics into account, they should in fact glow slightly, radiating as if they had a non-zero temperature. Nobody has ever seen this 'Hawking radiation'; it's far too faint. But Hawking's calculation has been thoroughly checked, and seems unassailable. The question is what to make of it. If black holes have a temperature, they can be studied using thermodynamics. In particular, their entropy, or disorder, can be calculated. Confirming Bekenstein's hope<sup>3</sup>, the entropy of a black hole turns out to be proportional to its surface area. Even better, the constant of proportionality can be worked out exactly. Now, the entropy of a system is effectively a measure of how many different states it can be in while looking the same on a large scale. For gas in a box, these different 'microstates' are easy to visualize: they are all the different ways that the atoms can wiggle about. But what are the microstates of a black hole?

This question has tortured physicists for decades, and dozens of answers have been suggested. As the entropy of a black hole is proportional to its surface area, a reasonable guess is that the microstates describe the geometry of its surface. This surface is not a physical object in the usual sense. Instead, it is an imaginary boundary, called the 'event horizon' — the closest distance that an object can approach a black hole before being inevitably sucked in. Still, in many ways the surface acts like a flexible membrane<sup>4</sup>. But to count the horizon's microstates (and get a sensible answer) requires a quantum mechanical description.

Some progress along these lines has come from 'loop quantum gravity'<sup>5</sup>. In this theory, the fabric of space is like a weave of tiny threads, and area comes in discrete units: each thread poking through a surface gives it a little bit of area<sup>6</sup> (Fig. 1). The surface area of a black hole, then, is generated by all the threads puncturing it. The event horizon is flat except at these punctures, where it can flex, and the microstates of the black hole are defined by the different ways the event horizon can flex in or out.

In 1997, Ashtekar and colleagues tried to calculate the entropy of a (non-rotating) black hole using this theory<sup>7</sup>. But in the process, they had to face up to a curious feature of loop quantum gravity. At the time, this theory was unable to predict the 'quantum of area' — the smallest amount of area carried by one of the threads from which space is woven. Without knowing this, they could check that the entropy of a black hole is proportional to its area, but they could not compute the constant of proportionality. But when life gives you lemons, sometimes you can make lemonade. After all, Hawking had already worked out this constant of proportionality by other means. So Ashtekar *et al.* were able to reverse



**Figure 1** The quantum black hole. In loop quantum gravity, space is envisaged as a fabric of woven threads. Where these threads puncture a surface, such as the event horizon of a black hole, they define its area.

their calculation and use Hawking's work to determine the quantum of area.

One year later, Hod<sup>8</sup> calculated the quantum of area in a completely different way. Instead of loop quantum gravity, he used a simple argument borrowed from Bohr's work on the hydrogen atom. He assumed that, like an atom, a black hole has a discrete spectrum of energy levels. Using computer simulations of the vibrational frequencies of black holes<sup>9</sup>, he guessed the spacing between these energy levels. For a non-rotating black hole, its energy determines its surface area, so from here Hod was able to work out the quantum of area. His result was 4.39444 times the Planck area (itself a puny unit of area, roughly  $10^{-70}$  m<sup>2</sup>). Finally, in a daring bit of numerology, he observed that this

number was very close to four times the natural logarithm of three, or  $4\ln(3)$ .

It would be nice to report that this result matched that of Ashtekar and colleagues. Unfortunately, it did not. But Dreyer<sup>1</sup> has now made a remarkable discovery. He has shown that if the loop-quantum-gravity calculation is modified slightly — in a way that is quite reasonable, but rather too technical to describe here — it gives a quantum of area equal to exactly  $4\ln(3)$  times the Planck area. When Dreyer posted his results on the preprint server late last year, the suspense among experts became almost unbearable. After all, Hod's observation relied on numerical calculations, and although his multiplicative factor of 4.39444 is certainly close to  $4\ln(3)$ , the next decimal place might fail to match it exactly. Fortunately, a new analysis by Luboš Motl<sup>10</sup> indicates that the match is, in fact, exact.

Could it all be just a coincidence, or has the quantum of area been discovered? Exciting as they are, these developments raise more questions than they answer. Nobody knows why the two calculations agree, or how to extend the calculations to the more general case of rotating black holes. Only time — or space — will tell.

John Baez is in the Department of Mathematics, University of California, Riverside, California 92521, USA.

e-mail: baez@math.ucr.edu

1. Dreyer, O. *Phys. Rev. Lett.* (in the press); preprint gr-qc/0211076 at <http://arXiv.org> (2002).
2. Hawking, S. *Commun. Math. Phys.* **43**, 199–220 (1975).
3. Bekenstein, J. *Phys. Rev. D* **7**, 2333–2346 (1973).
4. Thorne, K., Price, R. & Macdonald, D. (eds) *Black Holes: The Membrane Paradigm* (Yale Univ. Press, New Haven, 1986).
5. Rovelli, C. *Living Rev. Relativity* **1**, 1 (1998); [www.livingreviews.org/Articles/Volume1/1998-Irovelli](http://www.livingreviews.org/Articles/Volume1/1998-Irovelli)
6. Rovelli, C. & Smolin, S. *Nucl. Phys. B* **442**, 593–622 (1995).
7. Ashtekar, A., Baez, J., Corichi, C. & Krasnov, K. *Phys. Rev. Lett.* **80**, 904–907 (1998).
8. Hod, S. *Phys. Rev. Lett.* **81**, 4293–4296 (1998).
9. Nollert, H.-P. *Phys. Rev. D* **47**, 5253–5258 (1993).
10. Motl, L. preprint gr-qc/0212096 at <http://arXiv.org> (2002).

## Inflammation

# Border crossings

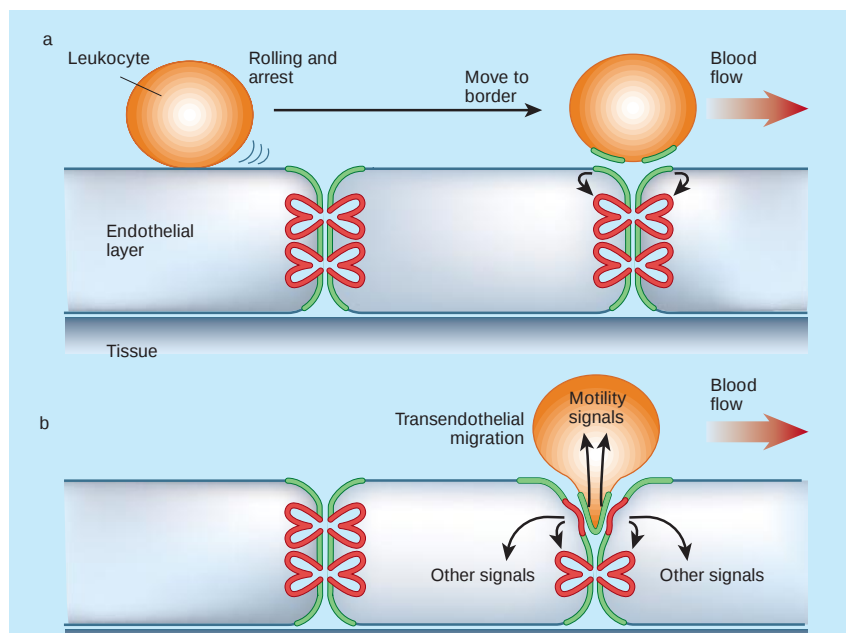
Ann Ager

**When inflammatory cells leave blood vessels to repair injured tissues, they are helped on their way by the endothelial cells lining the vessels. The invagination and expulsion of endothelial membrane may be the key.**

Inflammation — described in antiquity to comprise 'rubor' (redness), 'calor' (heat), 'tumor' (swelling) and 'dolor' (pain) — is the first response of tissues to injury. Its purpose is to recruit inflammatory cells and other blood components to an affected area, where they repair the damage. To reach a tissue in need, the inflammatory cells must squeeze their way out through the walls of blood vessels, and the cells that form the vessel walls are an active partner in this process.

But it is unclear how they accommodate and propel (or expel) migrating inflammatory cells. On page 748 of this issue, Mamdouh and colleagues<sup>1</sup> reveal surprising details of the mechanism.

Inflammatory cells are together known as white blood cells or leukocytes, and include neutrophils, lymphocytes, monocytes and mast cells. Normally, fluids and solutes are exchanged between blood and tissues at the smallest of the blood vessels, the



**Figure 1 Model for how inflammatory cells reach tissues in need.** To escape the blood and reach injured tissues, inflammatory cells (leukocytes) must cross the blood-vessel wall. **a**, Leukocytes move out of the blood flow and roll along the layer of endothelial cells that lines blood vessels; they then stop, and move to the junctions or borders between endothelial cells. These cells are held together in part through the self-binding protein PECAM on their surface (green); PECAM also occurs on leukocytes. Mamdouh *et al.*<sup>1</sup> show that PECAM is also located in membrane invaginations (red) in endothelial cells. **b**, When PECAM molecules on the endothelial-cell surface and on the leukocyte bind, the membrane invaginations are disgorged. This delivers stored PECAM to interact with leukocyte PECAM, perhaps generating signals that are required to move the leukocyte through the border. It may also plug the gap in the endothelial layer, preventing other blood components from leaking out. Curved arrows represent signalling events.

capillaries. But leukocytes instead emigrate from postcapillary venules, vessels that lead from capillaries into veins. The walls of venules have two layers: an inner, continuous lining of endothelial cells, and a surrounding mesh of support cells (pericytes). So leukocytes must crawl through both layers — in a process called diapedesis — in order to leave the blood at sites of injury. Details of how they do this, however, have been scant, as they leave little or no sign of their passage.

Some 30 years ago, it was shown that soluble proteins produced during inflammation cause endothelial cells to pull apart<sup>2</sup>, making the endothelial layer more permeable. The implication was that this could be the route that leukocytes and components of blood plasma take into inflamed tissues. Although it was not possible at the time to describe the precise anatomical path taken by the leukocytes, a previous analysis of tissue sections, showing the spatial relationship between the different cell types, suggested that leukocytes do squeeze out between adjacent endothelial cells<sup>3</sup>. Indeed, this is now well accepted. However, no obvious gaps in the blood-vessel lining are seen at sites of 'transendothelial' migration (transmigration)<sup>3</sup>. Moreover, there is a downside to a model in which leukocytes and plasma factors use the same exit strategy: there are likely to be situations

in which blood vessels need to be permeable to one or the other, but not to both.

So several research groups have been looking at how the permeability of the endothelial layer is regulated. There is a rapidly growing list of adhesion molecules that occur preferentially at the borders between adjacent endothelial cells<sup>4–6</sup>; these border proteins include VE-cadherin, junctional adhesion molecules 1, 2 and 3, PECAM (also known as CD31) and CD99. Some of them maintain the integrity of the endothelial layer<sup>7</sup>; all are capable of binding with like molecules and have the potential to stick adjacent cells together. Interestingly, all except VE-cadherin and junctional adhesion molecule 1 are also found on various types of leukocyte.

Mamdouh *et al.*<sup>1</sup> now build on these results, and suggest an unusual mechanism by which endothelial cells actively participate in leukocyte transmigration. Using cultured cells, and exploiting PECAM, VE-cadherin and CD99 as markers of endothelial-cell borders, the authors estimate that up to 30% of the membrane at these sites is invaginated in the absence of leukocytes. Border proteins move in and out of these invaginations, which are different from previously observed invaginations (such as recycling endosomes and caveolae), and may be a specialized property of endothelial cells.

Mamdouh *et al.* also find that, when in contact with a transmigrating leukocyte, an endothelial cell transiently increases its surface area by disgorging its invaginated membrane where the two cells meet. This process seems to require, among other things, binding of PECAM on the endothelial cell to PECAM on the leukocyte.

These intriguing findings support the idea that, even as a leukocyte forces its way out between endothelial cells, it plugs the gap, using the same adhesion molecules that endothelial cells use to stick together. The transient increase in contact area between leukocytes and endothelial cells demonstrated by Mamdouh *et al.* could serve to maintain the seal, preventing plasma factors from leaking out during the one or two minutes it takes for transendothelial migration. In a further refinement, engagement of border proteins by a leukocyte (rather than an adjacent endothelial cell) might generate the responses in both cell types that are needed to propel the leukocyte through the junction (Fig. 1). It is also tempting to speculate that cycles of membrane expulsion and invagination, which would lengthen and shorten the borders between endothelial cells like a purse-string, could actually propel leukocytes through.

The ability to rapidly mobilize a supply of membrane could also explain a previously puzzling alternative means of transmigration. Early studies<sup>8</sup> had concluded that neutrophils move between endothelial cells but that lymphocytes pass straight through them, actually penetrating the cell body. Subsequent work, however, showed that although lymphocytes do not pass through endothelial cells<sup>9</sup>, neutrophils can do so, albeit in very specific experimental conditions<sup>10</sup>.

The new results do not resolve this discrepancy, but do suggest a possible mechanism for 'transcellular' movement. Mamdouh *et al.* propose that a leukocyte could be engulfed by a single endothelial cell if invaginated membrane is discharged not onto the surface adjoining another endothelial cell, but instead onto the surface of the blood-vessel interior. (Indeed, some border proteins can become relocated to this surface, particularly during inflammation.) How exactly this might enable the leukocyte to pass through the endothelial cell is unclear, but it might exploit known cellular mechanisms for taking up and secreting molecules. Such a transcellular route could be particularly important in certain organs or for certain types of leukocyte. For example, in the central nervous system it would maintain the integrity of the endothelial layers that constitute the blood–brain barrier. Similarly, lymphocytes and monocytes continually survey the body for pathogens by moving from the bloodstream into tissues, and it might be advantageous for this to occur independently of changes in the permeability of the endothelium.

Although PECAM is not always involved in leukocyte transmigration<sup>11</sup>, Mamdouh and colleagues' model<sup>1</sup> will help in identifying the molecular events that regulate the endothelial lining during inflammation. Previous studies from the same group<sup>4</sup> provided compelling evidence that monocytes need PECAM to begin their penetration of endothelial layers, and CD99 to complete it. Other border proteins, such as the junctional adhesion molecules, might substitute for PECAM or CD99 during the transmigration of different types of leukocyte, depending on the tissue in need. It seems likely, then, that transendothelial migration involves a multi-step cascade of adhesive interactions that are coordinated in space and time, rather like the process by which leukocytes move out of the fast-flowing bloodstream and onto the inner endothelial surface before transmigration<sup>12</sup>.

Ann Ager is at the MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

e-mail: ann.ager@nimr.mrc.ac.uk

1. Mamdouh, Z., Chen, X., Pierini, L. M., Maxfield, F. R. & Muller, W. A. *Nature* **421**, 748–753 (2003).
2. Majno, G., Shea, S. M. & Leventhal, M. J. *Cell Biol.* **42**, 647–672 (1969).
3. Marchesi, V. T. Q. *J. Exp. Physiol.* **46**, 115–118 (1961).
4. Schenkel, A. R., Mamdouh, Z., Chen, X., Liebman, R. M. & Muller, W. A. *Nature Immunol.* **3**, 143–150 (2002).
5. Johnson-Leger, C., Aurand-Lions, M. & Imhof, B. A. *J. Cell Sci.* **113**, 921–933 (2000).
6. Lusinskas, F. W., Ma, S., Nusrat, A., Parkos, C. A. & Shaw, S. K. *Semin. Immunol.* **14**, 105–113 (2002).
7. Corada, M. et al. *Proc. Natl Acad. Sci. USA* **96**, 9815–9820 (1999).
8. Marchesi, V. T. & Gowans, J. L. *Proc. R. Soc. B* **159**, 283–290 (1964).
9. Schoeffl, G. I. *J. Exp. Med.* **136**, 568–588 (1972).
10. Feng, D., Nagy, J. A., Pyne, K., Dvorak, H. F. & Dvorak, A. M. *J. Exp. Med.* **187**, 903–915 (1998).
11. Duncan, G. S. et al. *J. Immunol.* **162**, 3022–3030 (1999).
12. Springer, T. A. *Cell* **76**, 301–314 (1994).

## Global change

# Dishing the dirt on coral reefs

Julia Cole

**Sediment is bad news for coral reefs. In Australia, large increases in sedimentation were a by-product of introducing intensive agricultural practices, and that may also apply to other parts of the world.**

Modern agriculture has transformed Earth's surface, and one consequence has been the increase in continental material that is eroded and transported to the oceans. There, it poses particular hazards to reef-building corals, by decreasing the availability of light and interfering with feeding. If sediment delivery to coral reefs can be evaluated and attributed to specific processes, then the effects of anthropogenic erosion on reefs can be quantified and managed more effectively. In a study described on page 727 of this issue, McCulloch and others<sup>1</sup> have cleverly developed a history of sedimentation on Australia's Great Barrier Reef by quizzing the corals themselves. Their results show that sedimentation on the reef increased dramatically following European settlement and agricultural expansion in northeast Australia.

Corals preserve a history of terrestrial sediment delivery because, as they build their skeletons from calcium and carbonate, trace amounts of other elements are incorporated. Suspended sediments in river water contain barium, which desorbs from these particles as the river water enters the ocean; during periods of high discharge, levels of the element rise in coastal waters<sup>2</sup>. As a coral grows, it inadvertently incorporates barium into its skeleton, in rough proportion to the ambient concentration. The skeletons grow at 1–2 centimetres per year, and annual bands allow ages to be assigned to the skeletal material. By



**Figure 1 Turbid time: sediment plume from the Burdekin river, Queensland, following Cyclone Joy in 1991.**

measuring the amount of barium (normalized to calcium) in the skeleton, McCulloch et al.<sup>1</sup> have developed an effective proxy record of terrestrial sedimentation on the reef that extends from AD 1750 to AD 1985. Their record shows that the influx of sediment onto the reef increased after 1870, shortly following European settlement. Both average and maximum values of barium are significantly higher after this time.

Could these changes in coral barium content be a consequence of other factors? Although such changes can reflect the upwelling of barium-rich deep waters<sup>3</sup>, this process is not significant at the inner-reef site analysed by McCulloch and colleagues. A climate-driven increase in the frequency and intensity of flooding, independent of

land-use change, might also cause variations in barium. But the authors use additional evidence from the coral skeletons (fluorescent banding, strontium content and oxygen isotopic content<sup>4,5</sup>) to show instead that, before 1870, flood waters still influenced the reef but contained much less barium. As domestic grazing and land clearance intensified after 1870, soil became more vulnerable to erosion during monsoon rains, and the barium content of river waters increased. From the change in the relationship between river discharge and barium calculated from coral data, the authors infer a five- to tenfold increase in the transport of suspended sediment to the reef following European settlement. This study is based on a single core, and reproducibility would add confidence to these quantitative estimates. But the basic conclusion that sedimentation has increased and become more variable is almost certainly robust.

Such results are not unique to the Great Barrier Reef. In East Africa, corals tell a similar tale of erosion exacerbated by the imposition of colonial agricultural practices in the early decades of the twentieth century. Unpublished data, collected by myself and colleagues on coral from Malindi Reef, Kenya, indicate a low and stable level of barium before about 1910 which rises dramatically by 1920, with a simultaneous increase in variance. This interval coincides with documented increases in soil erosion in Kenya during the early twentieth century<sup>6</sup>.

In both the Kenyan and Australian records, the increase of variance following land-use change is one of the most striking features. Recent maxima of barium reach two to three times the levels of eighteenth- and nineteenth-century maxima. The increased variance probably occurs because the soil is more easily eroded by intense rains under modern agricultural regimes, so a given amount of rainfall can mobilize more sediment (Fig. 1). These changes in variance are important for corals, which are likely to be sensitive to maximum sediment levels. In East Africa, modern sedimentation patterns appear to influence the distribution of reef species<sup>7</sup>. But we lack ecological data from before the transition in sedimentation regimes that would allow us to understand the full impact of these changes.

In both locations, increased river discharge is associated with extreme states of the El Niño–Southern Oscillation (ENSO) system. In Kenya, El Niño events bring the heaviest rains and greatest likelihood of flood, along with warmer seawater temperatures<sup>8</sup>. In northeastern Australia, major floods and increased likelihood of tropical storms are associated with La Niña events<sup>9</sup>. Human activity, in the form of changing land use, has added sedimentation to the list of stresses experienced by reefs during these ENSO extremes, making this inter-annual variability more problematic for the



corals. Moreover, it is possible that the ENSO system itself may be intensifying as a consequence of global warming<sup>10</sup>.

Another common feature of these records is that the increase in sedimentation occurs as a baseline shift around the time of agricultural intensification and not as a long-term, continuing trend. Is this because erosion has levelled off at the new, higher levels? Soil conservation measures and river damming could stabilize the amount of sediment transported to a reef. Alternatively, does the system that allows preservation of the barium signal in the coral saturate at some level, so that higher values cannot occur? Maybe the coral dramatically slows skeletal growth during times of highest sedimentation, so leaving little or no record. To understand these records better, more research is needed on how this tracer behaves as it moves from terrestrial sediments, to rivers and the oceans, and then into the coral skeleton.

Land-use intensification is widespread, so that many reefs close to continents or large islands are likely to have experienced increased delivery of sediment over the past century. As coastal populations increase, this phenomenon is likely to expand. Sedimentation is just one of many large-scale stresses threatening coral reefs<sup>11</sup>. Rising temperatures lead to bleaching, a response that can result in coral death, and increasing concentrations of atmospheric CO<sub>2</sub> make the

oceans' carbonate chemistry less favourable for calcification<sup>12</sup>.

Reef preservation efforts often focus on stemming the impact of locally significant threats, such as those from destructive fishing, mining and tourism. But mitigating large-scale reef stress from changing climate, ocean chemistry and land use will require regulation of the driving forces of global environmental change — which, thus far, nations have been reluctant to undertake. ■

Julia Cole is in the Department of Geosciences, Gould Simpson Building, 1040 East 4th Street, University of Arizona, Tucson, Arizona 85721, USA. e-mail: jcole@geo.arizona.edu

1. McCulloch, M. et al. *Nature* **421**, 727–730 (2003).
2. Shen, G. T. & Sanford, C. L. in *Global Ecological Consequences of the 1982–83 El Niño–Southern Oscillation* (ed. Glynn, P. W.) 255–284 (Elsevier, New York, 1990).
3. Lea, D. W., Boyle, E. A. & Shen, G. T. *Nature* **340**, 373–376 (1989).
4. McCulloch, M. T., Gagan, M. K., Mortimer, G. E., Chivas, A. R. & Isdale, P. *Geochim. Cosmochim. Acta* **58**, 2747–2754 (1994).
5. Isdale, P. J., Stewart, B. J. & Lough, J. M. *Holocene* **8**, 1–8 (1998).
6. Jacks, G. V. & Whyte, R. O. *Vanishing Lands: A World Survey of Soil Erosion* (Doubleday, New York, 1939).
7. McClanahan, T. R. & Obura, D. *J. Exp. Mar. Biol. Ecol.* **209**, 103–122 (1997).
8. Hastenrath, S., Nicklis, A. & Greischar, L. *J. Geophys. Res.* **98**, 20219–20235 (1993).
9. Lough, J. M. *Coral Reefs* **13**, 181–195 (1994).
10. Trenberth, K. E. & Hoar, T. W. *Geophys. Res. Lett.* **23**, 57–60 (1996).
11. Burke, L., Bryant, D., McManus, J. W. & Spalding, M. *Reefs at Risk: A Map-based Indicator of Threats to the World's Coral Reefs* (World Resources Inst., Washington DC, 1998).
12. Kleypas, J. A. et al. *Science* **284**, 118–120 (1999).

self-reactive T-helper 1 cells are involved<sup>4</sup>.

But although a fair amount of evidence implicates interleukin-12 in the development of autoimmune disease, the protein now seems to have an alibi. Unlike many other cytokines, which are the product of a single gene, interleukin-12 is composed of two proteins that are encoded by distinct genes. These two subunits are named p35 and p40, to reflect their relative molecular masses. The p35 subunit is continuously produced by macrophages and dendritic cells, but p40 is generated only when these cells encounter pathogens. The two proteins then combine to form the biologically active dimer, which is secreted and binds to target cells via a receptor that itself consists of two subunits, designated  $\beta 1$  and  $\beta 2$ . The targets of interleukin-12 are T cells, natural killer cells, dendritic cells and macrophages; in response to this cytokine, they release interferon- $\gamma$ .

This might seem fairly straightforward. The problem is that interleukin-12 is just one member of a small family of dimeric cytokines that regulate interferon- $\gamma$  production (Fig. 1). And, to confuse matters further, the p40 subunit is a component of both interleukin-12 and the newly discovered interleukin-23. The latter protein also has a p19 subunit<sup>5</sup> and has functions that are similar to, yet distinct from, those of interleukin-12 (reviewed in ref. 6). For instance, it too induces the production of interferon- $\gamma$  by T cells and dendritic cells<sup>7</sup>. Yet another complication is that the receptor for interleukin-23 comprises the  $\beta 1$  subunit of the interleukin-12 receptor, as well as a second protein.

And there's the rub: these similarities between interleukins 12 and 23 (and their receptors) call into question functions that have previously been attributed solely to interleukin-12. In the past, mice that were genetically engineered to lack p35 or p40 — and mice deficient in the  $\beta 1$  or  $\beta 2$  receptor proteins — were considered roughly equivalent, reflecting a lack of interleukin-12 activity. Abnormalities seen in people with p40 or  $\beta 1$  mutations were attributed to the same cause. But it is now clear that this simple view is inaccurate, a fact underscored by the finding that p40-deficient mice (which lack interleukins 12 and 23) are more immunocompromised than p35-deficient mice (which lack interleukin-12 alone). Furthermore, autoimmune disease has been seen<sup>8</sup> in mice that lack p35, but not in p40-deficient animals. So it is important to determine precisely what each cytokine does.

In their provocative study, Cua et al.<sup>1</sup> addressed this problem by producing p19-deficient mice — which lack interleukin-23, but not interleukin-12. The authors then compared the development of a 'model' of multiple sclerosis, experimental auto-

## Autoimmunity

# A case of mistaken identity

Wendy T. Watford and John J. O'Shea

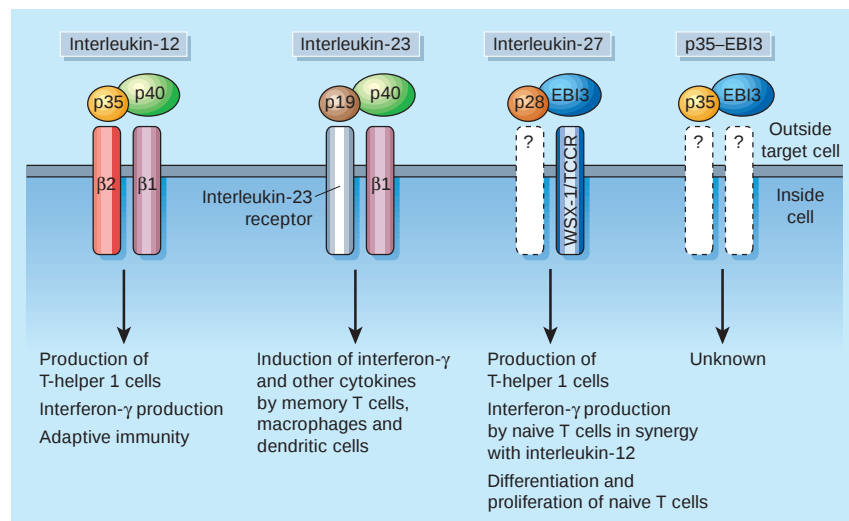
The interleukin-12 protein has been implicated in autoimmunity, but one complication is that it shares a subunit with a related protein. New work looks at the contribution of these proteins to autoimmunity in mice.

It is often quite a straightforward matter to attribute particular human diseases to the action of particular molecules. Sometimes, however, things are less clear cut. For instance, there seems to be considerable evidence that a protein known as interleukin-12 is a major contributor to autoimmune disorders. But on page 744 of this issue, Cua and colleagues<sup>1</sup> argue that this protein has been wrongly accused, at least in the case of a brain autoimmune disease in mice, and that the real culprit is a close relative.

Interleukin-12 is a cytokine — a member of a group of secreted proteins that have diverse roles in cellular differentiation and are especially important in regulating immune responses. Interleukin-12 itself is viewed as being a link between innate and adaptive immunity. The innate immune system consists of cells, such as macrophages and dendritic cells, that respond to generic features of intruders and act accordingly;

macrophages, for example, consume microorganisms. Macrophages and dendritic cells also release interleukin-12, and one of its effects is to drive the differentiation of naive T cells into T-helper type 1 cells<sup>2,3</sup>, which protect against many intracellular pathogens. T-helper 1 cells in turn produce another cytokine, interferon- $\gamma$ , which promotes responses by yet more immune cells that are adapted to the unique features of the pathogen in question.

Interleukin-12, then, is important for defence against intracellular pathogens. On the flip side, however, its ability to stimulate T-helper 1 cells and adaptive immunity has led to the proposal that it also actively contributes to several autoimmune diseases, including rheumatoid arthritis and inflammatory bowel disease. Moreover, abnormally high levels of interleukin-12 have been found in people with multiple sclerosis, a disease of the central nervous system in which



**Figure 1 The interleukin-12 family.** These cytokine proteins are released from macrophages and dendritic cells in response to pathogens, and bind to receptors on target cells, activating a variety of functions. In each cytokine a protein akin to p35 (such as p19 or p28) pairs with a soluble protein related to p40 (such as EBI3). p40 is a component of interleukins 12 and 23; their receptors both include the interleukin-12 receptor  $\beta 1$  subunit. Interleukin-27 comprises the p35 relative p28, and the p40 relative EBI3. It binds to the receptor subunit WSX-1/TCCR, and probably another, unknown, receptor subunit. EBI3 can also associate with p35, but the significance of this is unclear. Cua *et al.*<sup>1</sup> show that, because of subunit sharing between these proteins, interleukin-12 has been wrongly assumed to be a major contributor to autoimmune disease. Instead, interleukin-23 is the main culprit, at least in a brain autoimmune disease in mice.

immune encephalomyelitis (EAE), in p19-deficient and p35-deficient mice. The results show that although p19-deficient mice generate T-helper 1 cells and interferon- $\gamma$ , they do not develop EAE, but that administering interleukin-23 to these animals is enough to provoke the disease. In contrast, in p35-deficient mice the production of T-helper 1 cells is blocked (as expected, because these mice lack interleukin-12), but the animals are highly susceptible to EAE.

So interleukin-23 can trigger this autoimmune disease, but the role of interleukin-12 is more complicated. The authors find that adding interleukin-23 is not enough to induce EAE in p40-knockout mice (which lack both interleukins). But administration of interleukin-12 followed by interleukin-23 does cause disease. All of this suggests that interleukin-12 contributes to EAE, but that it also has a protective role in this particular model. Although this may seem strange, it is consistent with the finding that interferon- $\gamma$  also protects against EAE in an 'adjuvant' model<sup>9</sup>. Yet interferon- $\gamma$  often has a protective effect in adjuvant models, for reasons that are not clear. By contrast, the administration of interferon- $\gamma$  to people with multiple sclerosis exacerbates the disease<sup>9</sup>. Will the protective effects of interleukin-12 in mice be reflected in humans? The jury is still out.

Are there any other suspects to consider? Interleukin-12-related cytokines that share cytokine or receptor subunits continue to be identified, the newest member of the family

being interleukin-27. Composed of the EBI3 protein, which is related to p40, and a p28 protein, akin to p35, interleukin-27 promotes interferon- $\gamma$  production and drives the development of T-helper 1 cells. EBI3 may also be able to form dimers with other proteins, such as p35 itself. This might add another layer of complexity to interferon- $\gamma$  regulation. Moreover, cytokine receptors are notoriously promiscuous — exactly which receptors bind which cytokines needs to be sorted out. In short, there is still considerable work to be done to pin down the functions of cytokines in immunity and autoimmunity<sup>10</sup>.

What are the implications for treating patients? Antibodies that block interleukin-12 have been widely used in mouse models of autoimmune disease and are being tested in humans. It now seems that interleukin-23 should also be considered a target, although the antibodies against interleukin-12 might serendipitously block this cytokine, too. Treatments aimed at blocking protein subunits that are shared by interleukin-12 relatives (or their receptors) will also be of interest, as it is possible, given their common role in promoting interferon- $\gamma$  production, that many members of this family contribute to autoimmune disease. Certainly, there is far more complexity here than initially envisioned. But this gives ample opportunity for intervention. ■

Wendy T. Watford and John J. O'Shea are in the Molecular Immunology and Inflammation Branch, National Institutes of Arthritis, Musculoskeletal and Skin Disease, National Institutes of Health,



#### 100 YEARS AGO

In the essay to which he refers in his letter in *NATURE* of January 29, Dr. Wallace attaches less importance to the rearing of a few men of exceptional qualities than to the weeding out of the worst and raising the average; but surely, without giving undue and exclusive credit for advance to the pioneers and prophets, we may take it that men like Darwin and Wallace himself, to mention only one type, will, under natural selection, render the later more conscious steps of man's evolution easier. Dr. Wallace, in the letter referred to, speaks of the "fittest" not surviving under existing civilisation, meaning that many of the specialised types, which form important elements in our polymorphic communities, are not fittest to survive, and continue to reproduce their kind in more primitive or ideal communities. But this, of course, accords well with the principle of the "survival" of those types "fittest" to the actual environment. (Survival, of course, does not postulate direct reproduction any more than it postulates long life; the "worker" bees "survive.") Further, Dr. Wallace's hopeful attitude shows that he really trusts "natural selection" to steer the best races of man to a point whence their further, more self-conscious, progress (still, as always, under natural selection) will be more and more in accord with Nature's will, and so less wasteful and pain-fraught.

From *Nature* 12 February 1903.

#### 50 YEARS AGO

*Chinese Science Revisited (2).* By Dr. Joseph Needham. Another emphasis which must be mentioned is that on popular education. There is a touching and genuine thirst for scientific knowledge among the Chinese people. Shops which sell anatomical models and geological charts have a crowd around the windows all day. If the visitor from the West wanders into one of the modern bookshops on a Sunday, he will almost have to step over rows of children and boys and girls of various ages sitting on the floor and against the bookcases reading popular science, and not paying the slightest attention to him. The assistants never insist on readers buying books, though they usually do, as they are relatively extremely cheap ... Will it not make some difference to the world that five hundred million people are awakening to the significance of natural science and all that that implies?

From *Nature* 14 February 1953.

10 Center Drive MSC-1820, Bethesda, Maryland 20892-1820, USA.

e-mail: osheajo@mail.nih.gov

1. Cua, D. J. *et al.* *Nature* **421**, 744–748 (2003).
2. Trinchieri, G. *Nature Rev. Immunol.* **3**, 133–146 (2003).
3. Murphy, K. M. & Reiner, S. L. *Nature Rev. Immunol.* **2**, 933–944 (2002).
4. Gately, M. K. *et al.* *Annu. Rev. Immunol.* **16**, 495–521 (1998).
5. Oppmann, B. *et al.* *Immunity* **13**, 715–725 (2000).

6. Frucht, D. M. *Science STKE* <http://stke.sciencemag.org/cgi/content/full/sigtrans;2002/114/pe1> (2002).
7. Belladonna, M. L. *et al.* *J. Immunol.* **168**, 5448–5454 (2002).
8. Becher, B., Durell, B. G. & Noelle, R. J. *J. Clin. Invest.* **110**, 493–497 (2002).
9. Owens, T., Wekerle, H. & Antel, J. *Nature Med.* **7**, 161–166 (2001).
10. O'Shea, J. J., Ma, A. & Lipsky, P. *Nature Rev. Immunol.* **2**, 37–45 (2002).

## Astronomy

# Hot gas around the Galaxy

Amiel Sternberg

An extended system of highly ionized gas clouds that surrounds the Milky Way has been detected. This gas may be part of the original matter from which our Galaxy and its nearest neighbours formed.

How did galaxies form? Astronomical observations and theory suggest that the process began with the collapse of density perturbations in an originally smooth distribution of cosmic 'dark matter'. Baryonic matter, the ordinary visible matter that can be incorporated into stars, was then pulled into the growing condensations of dark matter. Eventually the baryons settled into the disks and elliptical structures observable today as galaxies. On the basis of data from the Far Ultraviolet Spectroscopic Explorer satellite (FUSE), Nicastro *et al.* (page 719 of this issue<sup>1</sup>) present evidence for the existence of a large reservoir of baryons around the Milky Way — which, they argue, may be a relic of the original matter that formed our Galaxy and its nearest neighbours.

The total baryon content of the Universe is a key cosmological parameter. Observations of the cosmic deuterium-to-hydrogen abundance ratio, a quantity that was fixed by nucleosynthesis in the hot Big Bang, imply that baryons constitute 4% of the total 'mass-energy' density of the Universe<sup>2</sup>. Remarkably, the same percentage is inferred from studies of the 'acoustic' oscillations in the primordial plasma, evident today as fluctuations in the cosmic microwave background radiation<sup>3</sup>.

Observations of intergalactic atomic hydrogen in clouds that existed when the Universe was about a tenth of its current age, the epoch when galaxy formation began, reveal a cosmologically expected baryon content. At that time, most of the baryons were distributed in a diffuse intergalactic medium, which was photoionized and heated to a temperature of around  $10^4$  K by the first stars and quasars. But in the present-day Universe, the baryon budget comes up short. Observations of stars, galaxies, clouds in the intergalactic medium and very hot ( $10^8$  K) X-ray-emitting plasma in galaxy clusters account for only a third of the expected cosmological baryon density<sup>4</sup>.

So where are the 'missing' baryons? One possibility is that they are contained in massive astrophysical compact halo objects (MACHOs), a collective term for objects such as isolated neutron stars, brown dwarfs or planetary masses that are faint and hard to detect. Alternatively, they could be hiding as dilute 'warm-hot' ionized clouds at temperatures of  $10^5$ – $10^7$  K. Such gas would be difficult to detect by conventional methods: it would be too highly ionized to allow it to be identified by observations of atomic hydrogen, yet too cool to be evident as thermally emitting X-ray plasma.

Hydrodynamic simulations<sup>5</sup> of galaxy formation suggest that much of the baryonic mass today could indeed be present in a warm-hot intergalactic medium (WHIM) at  $10^5$ – $10^7$  K. In this picture, WHIM is produced by the shock waves that inevitably occur as gas falls out of the much cooler, diffuse intergalactic medium and into the collapsing filamentary web of 'over-dense' regions where galaxies form. The shocked gas cannot cool efficiently because the WHIM densities are low, and as much as half of the primordial baryonic material is predicted to survive as WHIM, rather than condense into stars and galaxies.

Can the WHIM be detected? One possibility would be to detect the absorption-line signatures of highly ionized trace 'metals' (elements heavier than helium) that may be present<sup>6–8</sup>. The oxygen ions  $O^{5+}$ ,  $O^{6+}$  and  $O^{7+}$  — designated O VI, O VII and O VIII by astronomers — are particularly important candidates for this purpose. The available spectroscopic line transitions of the oxygen ions occur at ultraviolet and X-ray wavelengths, however, meaning that observations must be carried out above the Earth's atmosphere. This is now possible with the revolutionary spectroscopic capabilities of the FUSE satellite, and the Chandra and XMM-Newton orbiting X-ray observatories. Distant quasars serve as (point-like) sources of ultraviolet and X-ray radiation, and clouds are detected

as intervening oxygen-line absorbers along the lines of sight.

Indeed, one of the spectacular results of the FUSE mission has been the detection of many discrete O VI absorbers, distributed all over the sky<sup>9,10</sup>. They probably represent a heterogeneous collection of clouds that are at widely varying distances and have widely different origins. Some absorption probably originates in ionized gas close to, or within, the interstellar medium of the disk itself; some, however, may occur in clouds located well outside the Milky Way.

The 'radial velocities' of the O VI clouds, as determined by the Doppler shifts of the absorption lines, provide clues to their origin and location. Nicastro *et al.*<sup>1</sup>, and, independently, Sembach *et al.*<sup>10</sup>, have analysed high-velocity O VI absorbers, with absolute radial velocities in excess of  $100 \text{ km s}^{-1}$ , and as high as  $550 \text{ km s}^{-1}$ . Such velocities are larger than would be expected for motions confined to the Galactic disk. Nicastro *et al.* show that the kinematic properties of this ensemble may imply that most of the high-velocity O VI absorbers are distributed in a volume encompassing the entire Local Group of galaxies, consisting of the Milky Way, Andromeda and 30 or more less-massive galaxies. They note that an extended distribution of hot gas is consistent with very low gas densities inferred from X-ray absorption-line detections of O VII and O VIII. For such an extended distribution, the total mass of hot ionized gas is comparable to the dynamical mass of the Local Group, and the hot gas represents a major reservoir of baryons.

Is the extended distribution of high-velocity O VI absorbers the original WHIM? Nicastro *et al.* argue that it is. Given the uncertainties in the distance estimates, however, the clouds could also be reprocessed, metal-enriched gas that was ejected from the Milky Way (or its neighbours) by supernova explosions. Measurements of the metal abundances in the hot clouds will help to clarify this issue.

Amiel Sternberg is in the School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv 69978, Israel.

e-mail: amiel@wise.tau.ac.il

1. Nicastro, F. *et al.* *Nature* **421**, 719–721 (2003).
2. Scott, B., Nollet, K. M. & Turner, M. S. *Phys. Rev. D* **63**, 063512 (2001).
3. Goldstein, J. H. *et al.* *Astrophys. J.* (in the press); preprint astro-ph/0212517 at <<http://arXiv.org>> (2002).
4. Fukugita, M., Hogan, C. J. & Peebles, P. J. E. *Astrophys. J.* **503**, 518–530 (1998).
5. Davé, R. *et al.* *Astrophys. J.* **552**, 473–483 (2001).
6. Spitzer, L. Jr & Zabriskie, F. R. *Publ. Astron. Soc. Pacif.* **71**, 412–420 (1959).
7. Cen, R., Tripp, T. M., Ostriker, J. & Jenkins, E. B. *Astrophys. J.* **559**, L5–L8 (2001).
8. Tripp, T. M., Savage, B. D. & Jenkins, E. B. *Astrophys. J.* **534**, L1–L5 (2000).
9. Wakker, B. *et al.* *Astrophys. J. Suppl. Ser.* (in the press); preprint astro-ph/0208009 at <<http://arXiv.org>> (2002).
10. Sembach, K. R. *et al.* *Astrophys. J. Suppl. Ser.* (in the press); preprint astro-ph/0207562 at <<http://arXiv.org>> (2002).



## Plant science

## Self-help for the interactive root

*New Phytol.* **157**, 315–326 (2003)

Plant roots don't just remove substances from their surrounding environment — they put things back, too, in the form of mucilage. From their investigations, D. B. Read and colleagues conclude that mucilage modifies the physical and chemical properties of soil, making it easier for plants to take up water and nutrients.

Read *et al.* analysed the composition of mucilage from the roots of maize, wheat and lupin. They found the chemical content to be very different from that in the plant tissues, implying that the plants were not merely leaking but were exuding substances for a specific purpose. Mucilage contains phospholipids, which are surfactants and which, like a detergent, reduce the surface tension of water. The authors tested surfactant action on soil. They found that it should enable plants to take up water from smaller pockets in the soil than they could otherwise do, and that it promotes the dissolution of phosphorus from soil particles into the water, increasing the availability of this nutrient by 10%.

John Whitfield

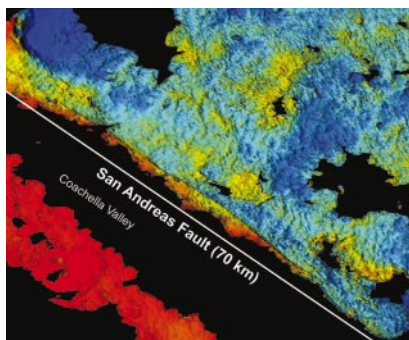
## Earth science

## Getting the measure of creep

*J. Geophys. Res.* **108**, doi:10.1029/2002JB001831 (2003)

**When creep occurs at earthquake faults, it releases stress in the uppermost crust that would otherwise build up and, if suddenly released, would manifest itself as an earthquake. As creep movement is typically less than 10 millimetres per year, its measurement requires high precision. This has been achieved at only a few faults, using creep-meters or dense networks of global-positioning-system receivers. Suzanne Lyons and David Sandwell show, however, that creep can be accurately followed by drawing up interferograms (see picture) from data collected by satellites equipped with synthetic aperture radar. They believe that it is even possible to measure fault movement in regions covered in vegetation or in built-up areas.**

Lyons and Sandwell used multiple satellite images of the southern San Andreas fault to automatically identify objects — such as buildings and outcrops of rock — that reflect radar consistently and so act as markers of creep. They further boosted sensitivity by layering the measurements to filter out transient phenomena such as turbulence in the atmosphere. The authors claim that the method can monitor creep over large areas



**Displacement in southern California.** Red and blue, respectively, show motion away from and towards the satellite. Regions without good point 'scatterers' are black.

and at higher spatial resolution than can the global-positioning system. **Tom Clarke**

## Genomics

## A bumper harvest of potato genes

*Plant Physiol.* **131**, 419–429 (2003)

The common potato (*Solanum tuberosum*) is the world's fourth largest food crop, with some 300 million tonnes grown annually. A member of the family Solanaceae, which also includes plants such as deadly nightshade and tomato, the potato has many similarities to its relatives. But it also has some outlandish characteristics, notably its tubers — the starch-filled root structures that are so versatile in cooking. C. Robin Buell and colleagues have now used genomics to investigate tuber development.

The group sequenced some 83,000 complementary DNAs from potatoes, representing nearly 20,000 genes; cDNAs provide an indication of which genes are active in a given tissue at a given time. The cDNAs came from four stages of tuber development: stolon, microtuber, dormant tuber and sprouting eyes. The authors thereby identified genes with a possible role in tuber development.

Potatoes are also prey to a notorious disease called potato late blight. By comparing cDNAs from uninfected and blight-infected leaves, and from plants either resistant or susceptible to blight, the group could pinpoint genes involved in resistance as well.

Christopher Surridge

## Particle physics

## Photon charge reaches new limits

*Phys. Rev. D* **67**, 017701 (2003)

**Photons have no electrical charge, according to conventional wisdom. But are we sure? A new measurement places an upper limit on the charge a photon might carry of**

$8.5 \times 10^{-17}$  times the charge on an electron ( $e$ ). That is only a slight improvement on the previous limit of  $10^{-16}e$  from a lab measurement, but Y. K. Semertzidis and colleagues claim that their experiment could, with minor adjustments, achieve a sensitivity of  $10^{-21}e$ .

Astronomical observations of radiofrequency photons from a millisecond pulsar have previously placed a far more stringent constraint on the photon's charge:  $5 \times 10^{-30}e$ . But this figure relies on uncertain assumptions about the nature of intergalactic magnetic fields.

The new experiment is simple in principle. A laser beam is passed through a magnet and, after travelling 21 metres, is focused onto a photodiode. The magnet's field is modulated, and charged photons should exhibit this same modulation frequency, in phase with the magnet, as they arrive at the photodiode. The researchers see no sign of this modulation in the (noisy) photodiode signal. Using a longer flight path and faster modulation could increase the sensitivity by up to five orders of magnitude, the authors say.

Philip Ball

## Cell biology

## Double trouble with mutant p53

*Mol. Cell* **11**, doi:10.1016/S1097276503000509 (2003)

In the event of a cellular emergency such as stress or DNA damage, the p53 protein can trigger apoptosis — the tidy demolition of the cell. This helps to prevent the uncontrolled cell proliferation that is a hallmark of cancer. Being a transcription factor, p53 is known to activate genes that are involved in apoptosis. But it has long been suspected of also having a more direct role in the cell's demise. Motohiro Mihara *et al.* now place p53 at the heart of the action — the cell's energy-producing organelles, the mitochondria.

Gene activation by p53 results in the perforation of mitochondrial membranes, the consequent leakage of apoptotic proteins, activation of specific enzymes and, ultimately, the shredding of proteins and DNA. Exactly how the mitochondrial membranes are perforated is not known, but the extended family of Bcl-2 proteins is intimately involved.

Mihara *et al.* now find that p53 binds two minders of mitochondria, Bcl-xL and Bcl-2. This prevents them from protecting mitochondria and results in the release of apoptotic proteins. Interestingly, tumour-derived p53 mutants that cannot bind DNA and stimulate transcription seem also to be unable to bind Bcl-xL. So, by blocking two routes to cell death, one mutation may mean double trouble.

Marie-Thérèse Heemels

# Adult persistence of head-turning asymmetry

A neonatal right-side preference makes a surprising romantic reappearance later in life.

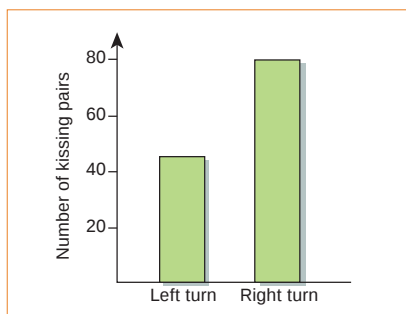
A preference in humans for turning the head to the right, rather than to the left, during the final weeks of gestation and for the first six months after birth<sup>1,2</sup> constitutes one of the earliest examples of behavioural asymmetry and is thought to influence the subsequent development of perceptual and motor preferences by increasing visual orientation to the right side<sup>3,4</sup>. Here I show that twice as many adults turn their heads to the right as to the left when kissing, indicating that this head-motor bias persists into adulthood. My finding may be linked to other forms of sidedness (for example, favouring the right foot, ear or eye) that do not become established until long after the newborn head-turning preference has disappeared<sup>5,6</sup>.

I observed kissing couples in public places (international airports, large railway stations, beaches and parks) in the United States, Germany and Turkey. The head-turning behaviour of each couple was recorded for a single kiss, with only the first being counted in instances of multiple kissing. The following criteria had to be met to qualify: lip contact, face-to-face positioning, no hand-held objects (as these might induce a side preference), and an obvious head-turning direction during kissing. Subjects' ages ranged from about 13–70 years.

Of 124 kissing pairs, 80 (64.5%) turned their heads to the right and 44 (35.5%) turned to the left (Fig. 1). This roughly 2:1 ratio is significantly different from 50% ( $\chi^2 = 5.34$ , d.f. = 1,  $P < 0.05$ ). As the couples come from a biologically adult age range, this result indicates that adults have a head-turning bias towards the right side, just like embryos and newborns.

Preferential use of the right foot, ear or eye is also evident as a 2:1 ratio<sup>7</sup>, which raises the possibility that these biases may be decreed by the observed head-turning preference. The incidence of right-handedness, however, is about 8:1 (ref. 8), so this particular asymmetry cannot be the result of a simple bias attributable to a right-sided head-turning tendency — the genetic origins of this trait may be different<sup>8</sup>, or cultural factors may have modified an original 2:1 pattern.

It takes two people to kiss (Fig. 2), so the bias of individuals cannot be judged from observations made on pairs. For example, what happens when a right-turner kisses a left-turner? If we assume that kissers with opposite biases could go either way with equal probability, then the individual biases should match those of the couples. For



**Figure 1** The number of couples who turn their heads to the right rather than to the left when kissing predominates by almost 2:1 (64.5%: 35.5%;  $n = 124$  couples).

example, if the individual bias is also 2:1 to the right and if couplings are random, then four of nine pairs would be right kissers, one of nine would be left-kissing, and four of nine would be mixed; if choice is random in the last group (that is, two of these four pairs are right-turning and two are left-turning), the result for the nine couples would be a right-turn kissing bias in six of them. This presumed 2:1 distribution for individuals thus predicts the observed 2:1 ratio for couples, indicating that the group bias may also reflect individual asymmetry.

In birds, a preference for turning the head to the right before hatching induces motor<sup>9</sup>, visual<sup>10</sup> and cognitive<sup>11</sup> asymmetries. If a similar effect occurs in humans and an initial asymmetry in the direction of head turning stimulates various functional left–right differences<sup>12</sup>, then the head-turning bias of the newborn would have to overlap in time with the establishment of adult asymmetries. My finding that turning of the head to the right is preferred throughout adulthood suggests that, by fostering a constant bias towards the right, this mechanism may be able to induce or enhance right-sided asymmetries of perception and action.

**Onur Güntürkün**

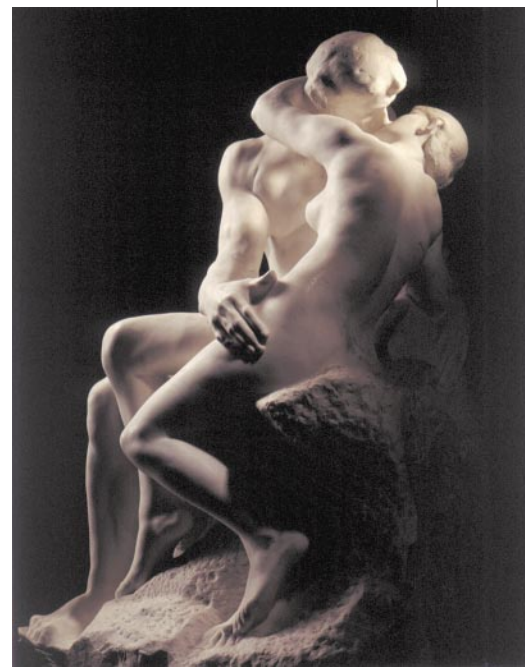
*Biopsychologie, Fakultät für Psychologie, Ruhr-*

COMMUNICATIONS ARISING

**Planetary science**

## Volcanism or aqueous alteration on Mars?

The Mars Global Surveyor's Thermal Emission Spectrometer has identified two global spectral surface types in martian dark regions, which were initially interpreted as being of basaltic (type 1) and intermediate (type 2, basaltic andesite to andesite) volcanic composition<sup>1</sup>. Wyatt and



**Figure 2** *The Kiss*: the couple in Auguste Rodin's masterpiece are turning their heads to the right to kiss each other.

Universität Bochum, 44780 Bochum, Germany  
e-mail: onur.guentuerkuen@ruhr-uni-bochum.de

- Ververs, I. A. P., de Vries, J. I. P., van Geijn, H. P. & Hopkins, B. *Early Hum. Dev.* **39**, 83–91 (1994).
- Konishi, Y., Mikawa, H. & Suzuki, J. *Dev. Med. Child Neurol.* **28**, 450–457 (1986).
- Konishi, Y., Kuriyama, M., Mikawa, H. & Suzuki, J. *Dev. Med. Child Neurol.* **29**, 751–757 (1987).
- Coryell, J. F. & Michel, G. F. *Infant Behav. Dev.* **1**, 245–257 (1978).
- Michel, G. F., Sheu, C.-F. & Brumley, M. R. *Dev. Psychobiol.* **40**, 1–13 (2002).
- Gentry, V. & Gabbard, C. J. *Gen. Psychol.* **122**, 37–45 (1995).
- Reiss, M. & Reiss, G. *Percept. Mot. Skills* **85**, 569–574 (1997).
- Corballis, M. C. *Psychol. Rev.* **104**, 714–727 (1997).
- Casey, M. B. & Martino, C. M. *Dev. Psychobiol.* **37**, 13–24 (2000).
- Skiba, M., Diekamp, B. & Güntürkün, O. *Behav. Brain Res.* **134**, 149–156 (2002).
- Rogers, L. J. & Andrews, R. J. (eds) *Comparative Vertebrate Lateralization* (Cambridge Univ. Press, Cambridge, UK, 2002).
- Nicholls, M. E., Clode, D., Wood, S. J. & Wood, A. G. *Proc. R. Soc. Lond. B* **266**, 1517–1522 (1999).

**Competing financial interests:** declared none.

McSween<sup>2</sup> suggest that northern-hemisphere occurrences of the type-2 spectrum are instead representative of weathered basalt, on the basis of ambiguity in the spectral interpretation and overlap between the distribution of this material and the outline of a putative ancient ocean. Although the spectral data may be open to interpretation within the limits of current understanding, creating weathered basalt and explaining its distribution is problematic. These competing hypotheses have significantly different implications for the igneous and aqueous

history of Mars, and it is important to continue debating the merits of each.

The model of the type-2 spectrum with the lowest r.m.s. error includes both high-silica glasses (silica) and phyllosilicates (clay minerals), in which the silica was first interpreted as a primary igneous phase<sup>1</sup>. There is no evidence in the spectral or geological data from Mars that justifies *a priori* removal of silica from such models. Martian dark-region materials are primarily active sands<sup>1,3</sup>, and so are constantly abraded and are unlikely to develop or retain rinds or coatings. Infrared spectra of terrestrial sands derived from weathered basalts do not show mineralogical evidence of weathering<sup>4</sup>. Visible-spectrum and near-infrared data have not identified abundant phyllosilicates in martian dark regions<sup>5</sup>.

Nonetheless, Wyatt and McSween exclude silica from their model, forcing it instead to include large quantities of phyllosilicates<sup>2</sup>. The increased abundance of phyllosilicate is attributed to the weathering of pyroxenes, but pyroxenes should also alter to amphibole in the presence of water, and there is no evidence of any amphibole in martian spectra<sup>6</sup>. (The lack of amphibole is not inconsistent with intermediate materials<sup>7</sup>.) K-feldspar minerals in the model are explained as the result of weathering of basalt, but a source of potassium for producing K-feldspar is not apparent in the protolith type-1 basalt, and K-feldspar is typically a product of hydrothermal, rather than ambient, aqueous alteration of basalt<sup>8</sup>. Although a reasonable fit to the data can be obtained without silica, this cannot be used to suggest that silica is not present, because it was intentionally excluded.

Columbia River basalt is the proposed martian analogue<sup>2</sup>. The mineralogy of this rock is similar to that of the type-2 surface<sup>1</sup>, including the presence of silica on both unweathered and weathered surfaces<sup>2</sup>. However, the Columbia River basalt rock has not been subaqueously altered and is not a realistic analogue for low-temperature subaqueous alteration. Silica can be produced by hydrothermal alteration of basalt, but is unlikely to form beneath a martian ocean. The physical character of the sample (rock) is unrealistic for comparison to Mars, where active sands that are unlikely to have coatings or rinds dominate the surfaces in question.

Type-2 spectra characterize extensive areas of Mars, including the southern highlands<sup>1</sup>, and are present in vast contiguous areas up to 1,000 km outside and 2 km higher in elevation than the putative “shoreline” boundaries<sup>1</sup>. Wyatt and McSween dismiss these incongruities and do not explain them<sup>2</sup>; no explanation is provided either as to how the Pathfinder landing-site analyses fit into the weathered-basalt hypothesis<sup>2</sup>. Some interpretations of Pathfinder data identify these as andesitic



Mars, seen from the Mars Global Surveyor orbiter. The distribution of surface rock types gives clues to the planet's geological evolution.

materials<sup>9</sup>, yet they are located in a region that is contiguous with ‘weathered basalt’.

In judging working hypotheses, it is common to consider the complexity of the models required to explain the data. Reconciling the global distribution of type-2 materials with an aqueous-weathering hypothesis requires a complicated model in which all type 2 materials within the ‘shoreline’ are weathered basalt and in which one or more processes (other than subaqueous alteration under an ocean) produced material with identical spectra just across the shoreline and over the rest of the planet.

Terrestrial results may provide a viable mechanism for producing large quantities of intermediate volcanics<sup>10</sup>; such mechanisms have not been evaluated for Mars, making it premature to dismiss production of significant amounts of intermediate compositions. The presence of slightly weathered intermediate volcanics<sup>1</sup> is a simpler hypothesis that explains all of the type-2 observations without invoking the existence of an ancient ocean.

**Victoria E. Hamilton\*, Philip R. Christensen, Joshua L. Bandfield**

Department of Geological Sciences, Arizona State University, Tempe, Arizona 85287-6305, USA

\*Present address: Hawaii Institute of Geophysics and Planetology, University of Hawaii, Honolulu, Hawaii 96826, USA

e-mail: hamilton@higp.hawaii.edu

1. Bandfield, J. L., Hamilton, V. E. & Christensen, P. R. *Science* **287**, 1626–1630 (2000).
2. Wyatt, M. B. & McSween, H. Y. Jr *Nature* **417**, 263–266 (2002).
3. Christensen, P. R. & Moore, H. J. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 686–729 (Univ. Arizona Press, Tucson, 1992).
4. Bandfield, J. L., Edgett, K. S. & Christensen, P. R. *J. Geophys. Res.* **107**, doi:10.1029/2000JE001469 (2003).
5. Bell, J. F. III in *Mineral Spectroscopy: A Tribute to Roger G. Burns* (eds Dyar, M. D., McCammon, C. & Schaefer, M. W.) 359–380 (Geochem. Soc., Houston, 1996).
6. Bandfield, J. L. *J. Geophys. Res.* **107**, 10.1029/2001JE001510 (2002).
7. Wyatt, M. B., Hamilton, V. E., McSween, H. Y., Christensen, P. R. & Taylor, L. A., *J. Geophys. Res.* **106**, 14711–14732 (2001).
8. Thompson, G. in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A. et al.) 225–278 (Plenum, New York, 1983).
9. Foley, C. N., Economou, T. E., Dietrich, W. & Clayton, R. N. *Meteor. Planet. Sci.* **35**, 55–56 (2000).
10. Marsh, B. D. *Geochim. Cosmochim. Acta* **66**, 2211–2229 (2002).

Wyatt and McSween reply — Multiple working hypotheses are prudent in the absence of definitive answers, and we suggest that there is no definitive interpretation of the surface type-2 spectrum from the Mars Global Surveyor's Thermal Emission Spectrometer (TES)<sup>1,2</sup>.

Demonstrations that some phyllosilicates (clay minerals) and oxidized basaltic glasses are spectrally similar to high-silica volcanic glasses<sup>1,3,4</sup> are critical because abundant high-silica glass is an effective parameter for distinguishing basalt from andesite<sup>5</sup> and has been used for petrological classification of TES surface type-1 and type-2 spectra. We excluded high-silica glass to examine one possible end-member of modelled mineralogies for surface type 2. Similarly, spectral end-members of oxidized and altered glasses were not included in studies by Bandfield *et al.*<sup>2</sup> and Hamilton *et al.*<sup>6</sup>; spectral comparison of these phases with igneous glasses was therefore not evaluated. Our preferred interpretation of type-2 spectra involves a combination of igneous minerals with alteration glass and clays.

We modelled 31% (by volume) clay and 14% K-feldspar for surface type 2, whereas Bandfield *et al.*<sup>2</sup> modelled 15% “sheet silicate” and 6% K-feldspar. The K-feldspar in both models<sup>1,2</sup> and “sheet silicate” in Bandfield *et al.*<sup>2</sup> are at or below the TES's 10–15% detectability limit. We suggested that clays and minor K-feldspar could be low-temperature aqueous-alteration products of pyroxenes and feldspars<sup>7</sup>. Pyroxenes in volcanic rocks can also alter to amphiboles in the presence of water, but such alteration is usually associated with high-temperature magmatic fluids or metamorphism. Hornblende is, however, a common phase in andesite, and the absence of modelled amphibole may be an issue in the intermediate-volcanic interpretation.

The argument of terrestrial sands derived from weathered basalts not exhibiting mineralogical evidence of weathering<sup>8</sup> depends greatly on the extent of alteration, crystallinity, distance travelled and depositional environment. Moreover, the martian northern plains may contain a mixture of unaltered and thoroughly altered basaltic sands.

We inferred from the correlation between the locations of altered basalt on Mars and a putative northern ocean basin that lowland plains could be composed either of basalts that are partly weathered under submarine conditions, or of weathered basaltic sediments transported into this depocentre. Our Columbia River basalt sample indicated that a single rock can exhibit the type of variability seen in type-1 and type-2 materials — it did not represent ‘Mars in a hand sample’ and is not representative of subaqueously altered basalts. Similarly, in Bandfield *et al.*<sup>2</sup> the terrestrial analogue of the type-2 surface is



a subduction-related andesite that was not used to promote plate tectonics on Mars.

The classification of “andesite” rocks at the Mars Pathfinder landing site is tenuous: this name (actually icelandite) was originally assigned on the basis of its major-element chemistry, noting that a sedimentary origin or weathering rind could not be ruled out without textural or mineralogical data<sup>9</sup>. A re-analysis of  $\alpha$ -proton X-ray spectrometer chemistry<sup>10</sup> indicates that Pathfinder rocks may have a high water content, which supports a non-igneous classification.

Formation mechanisms on a global scale are required to explain the extensive distribution of type-2 materials in the northern lowlands. The weathered-basalt interpretation would predict a large body of water or sedimentary depocentre in the northern lowlands, but does not account for local occurrences of type-2 materials in the southern highlands. These might be either andesites formed by igneous fractionation or the result of local weathering processes.

The production of fractionated magmas of intermediate composition is an inefficient process unless it is promoted by dissolved water, so the occurrence of vast amounts of andesite would probably require a wet source region and efficient transport of less dense magmas. The intermediate-volcanism interpretation requires large-scale melting of the thin northern crust and flooding with high-silica volcanics, as well as isolated melting pockets in the thick and ancient southern highlands.

However, the mechanism for rejuvenated mantle melting, the degree of magma fractionation and crustal assimilation, and the creation of local hotspots are all unresolved issues. Furthermore, there is a poor correlation between the transition of surface type-1 and type-2 materials with the transition from thick to thin crust, which suggests that the distribution of materials is a result of surficial processes rather than crust–mantle ones.

**Michael B. Wyatt\*, Harry Y. McSweeney Jr**

Department of Geological Sciences, University of Tennessee, Knoxville, Tennessee 37996-1410, USA

\*Present address: Department of Geological Sciences, Arizona State University, Tempe, Arizona 85287-6305, USA

e-mail: michael.wyatt@asu.edu

- Wyatt, M. B. & McSweeney, H. Y. Jr *Nature* **417**, 263–266 (2002).
- Bandfield, J. L., Hamilton, V. E. & Christensen, P. R. *Science* **287**, 1626–1630 (2000).
- Minitti, M. E., Rutherford, M. J. & Mustard, J. F. *LPSC* **31**, CD #1282 (2000).
- Noble, S. K. & Pieters, C. M. *LPSC* **32**, CD #1230 (2001).
- Wyatt, M. B., Hamilton, V. E., McSweeney, H. Y., Christensen, P. R. & Taylor, L. A. *J. Geophys. Res.* **106**, 14711–14732 (2001).
- Hamilton, V. E., Wyatt, M. B., McSweeney, H. Y. Jr & Christensen, P. R. *J. Geophys. Res.* **106**, 14733–14746 (2002).
- Thompson, G. in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A., Bostrom, K., Laubier, L. & Smith, K. L. Jr) 225–278 (Plenum, New York, 1983).
- Bandfield, J. L., Edgett, K. S. & Christensen, P. R. *J. Geophys. Res.* (in press).
- McSweeney, H. Y. Jr *et al. J. Geophys. Res.* **104**, 8679–8715 (1999).
- Foley, C. N., Economou, T. E. & Clayton, R. N. *LPSC* **32**, CD #1979 (2001).

COMMUNICATIONS ARISING

Physiology

## Why does metabolic rate scale with body size?

A long-standing problem has been the origin of quarter-power allometric scaling laws that relate many characteristics of organisms to their body mass<sup>1,2</sup> — specifically, whole-organism metabolic rate,  $B = aM^b$ , where  $M$  is body mass,  $a$  is a taxon-dependent normalization, and  $b \approx 3/4$  for animals and plants. Darveau *et al.*<sup>3</sup> propose a multiple-cause model for mammalian metabolic rate as the “sum of multiple contributors”,  $B_i$ , which they assume to scale as  $B_i = a_i M^{b_i}$ , and obtain  $b \approx 0.78$  for the basal and 0.86 for the maximally active rate,  $\dot{V}_{O_2}^{\max}$ . We argue, however, that this scaling equation is based on technical, theoretical and conceptual errors, including misrepresentations of our published results<sup>4,5</sup>.

All of the results of Darveau *et al.*<sup>3</sup> follow from their equation (2),  $B = a \sum c_i M^{b_i}$ , which they neither derive nor prove. As control coefficients<sup>6</sup>,  $c_i$ , and exponents,  $b_i$ , are dimensionless, this must be incorrect because it violates dimensional homogeneity, leading to different results for  $b$  that depend on the units of mass: for the basal rate,  $b \approx 0.76$  when  $M$  is in kilograms, and  $b \approx 1.08$  when  $M$  is in picograms.

Now, by definition,  $c_i \equiv \partial \ln B / \partial \ln B_i$ , which leads to the standard sum rules<sup>6</sup>  $\sum c_i = 1$  and  $\sum c_i \varepsilon_i = 0$ , where  $\varepsilon_i = b - b_i$  with  $b(M) \equiv d \ln B / d \ln M$ , the slope of  $\ln B$  versus  $\ln M$ , and  $b_i(M) \equiv d \ln B_i / d \ln M$ . This gives  $b = \sum c_i b_i$ , the equations that Darveau *et al.* should have used to determine  $b$  from the empirical  $c_i$  and  $b_i$ . These formulae are very general, requiring no assumptions about how  $B$  and  $B_i$  scale, or whether the  $B_i$  are connected in parallel or in series. Darveau *et al.*, however, use  $B = \sum B_i$ , implying that the  $B_i$  are added in parallel and, as such, their model is simply a consistency check on the conservation of energy, which requires all “ATP-utilizing processes”<sup>3</sup> (in parallel) to sum to  $B$  and so must be trivially correct. This gives  $c_i = (a_i/a) M^{-\varepsilon_i}$  and  $B = a \sum (c_i M^{\varepsilon_i}) M^{b_i}$ , which is the (dimensionally) correct version of equation (2).

As Darveau *et al.* take  $a$  and  $a_i$  as constant, their  $c_i$  must scale as  $M^{-\varepsilon_i}$ . However, they assume that  $c_i$  (and  $b_i$ )  $\propto M^0$ , which requires  $b$  (which equals  $\sum c_i b_i$ )  $\propto M^0$ , thereby contradicting their equation (2), in which  $b$  depends on  $M$ . This inconsistency in the  $M$ -dependence of  $b$  is concealed in their plots, which cover only three orders of magnitude in  $M$ , over which  $b$  is almost constant (about 0.78 for basal). However, when their analysis is extended to the realistic eight orders of magnitude spanned by mammals, their  $b$  increases with  $M$  to an average value of about 0.85 and, for  $\dot{V}_{O_2}^{\max}$ , to about 0.98, which are

both inconsistent with other data<sup>1,2</sup>.

Darveau *et al.* have taken their value for  $b_i$  from empirical data, without explaining why  $B$ , or  $B_i$ , scales nonlinearly with  $M$ , or why most  $b_i \approx 3/4$ . Understanding these features is the real challenge — the formulation of Darveau *et al.* is therefore hardly fundamental. By contrast, our theory<sup>4,5</sup> is grounded in basic principles of geometry, physics and biology, and offers a general unifying explanation for these and the other quarter-power scalings that are so pervasive in biology.

**Geoffrey B. West\*†, Van M. Savage\*†, James Gillooly‡, Brian J. Enquist§, William H. Woodruff\*†, James H. Brown\*††**

\*Los Alamos National Laboratory, Los Alamos,

New Mexico 87545, USA

e-mail: gbw@lanl.gov

†The Santa Fe Institute, Santa Fe,

New Mexico 87501, USA

‡Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA

§Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

- Schmidt-Nielsen, K. *Scaling: Why is Animal Size So Important?* (Cambridge Univ. Press, Cambridge, UK, 1984).
- Peters, R. H. *The Ecological Implications of Body Size* (Cambridge Univ. Press, Cambridge, UK, 1983).
- Darveau, C.-A., Suarez, R. K., Andrews, R. D. & Hochachka, P. W. *Nature* **417**, 166–170 (2002).
- West, G. B., Brown, J. H. & Enquist, B. J. *Science* **276**, 122–126 (1997).
- West, G. B., Brown, J. H. & Woodruff, W. H. *Proc. Natl Acad. Sci.* **99**, 2473–2478 (2002).
- Westerhoff, H. V. & Dam, K. *Thermodynamics and Control of Biological Free-Energy Transduction* (Elsevier, Amsterdam, 1987).

Physiology

## Allometric cascades

An allometric power-law relationship between metabolic rate and the mass of living organisms has been observed over many orders of magnitude in mass, indicating that (among other things) a characteristic mass scale is not applicable. Darveau *et al.*<sup>1</sup> present a multiple-cause cascade model of metabolic allometry, which has been hailed as a new perspective on comparative integrative physiology<sup>2</sup> and scaling relationships<sup>3</sup>. Here we show that this cascade model is flawed and is therefore meaningless both for control of metabolic rate in an organism of a given size and for scaling of the metabolic rate.

The basic equation of the cascade model<sup>1</sup> is

$$MR = a \sum_i c_i M^{b_i} \quad (1)$$

where MR is the metabolic rate in any given state,  $M$  is body mass,  $a$  is a coefficient,  $b_i$  is the scaling exponent of the process  $i$ , and  $c_i$  is the control coefficient of the process  $i$ . The control coefficients are chosen so that

$$\sum_i c_i = 1 \quad (2)$$

The justification for the model is that the allometric cascade arises from the layering of function at various levels of organization<sup>1</sup>, with the numerous steps involved in pathways of demand and supply each characterized by their own  $b_i$  and  $c_i$  values. Differences in the scaling of the basal metabolic rate and the maximum metabolic rate are then accounted for by noting that the  $c_i$  coefficients are different in the two cases and thus the calculation involves mixing different 'cocktails' of the components, as was assumed in ref. 4.

Dimensional analysis shows that the units of  $c_i$  depend crucially on the exponent  $b_i$ . This follows on noting that, because the units of MR and  $a$  are fixed, the units of  $c_i M^{b_i}$  are independent of  $i$ . Thus the requirement in equation (2) that the  $c_i$  values add up to 1 is erroneous because it is meaningless to add quantities with different units and require that the sum be unity.

Furthermore, the model (equation (1)) critically depends on the units of mass that are chosen. Consider a toy example of a two-process system with  $c_1$  and  $c_2$  equal to 1/2 (in different units, as above) and the exponents  $b_1$  and  $b_2$  being 1 and 2, respectively, with a body mass of 1 kg. (Note that this point does not depend on the choice of these numbers.) When body mass is measured in kilograms, the contributions of the two processes to  $MR/a$  are equal. However, if the mass were measured in grams, the second process would contribute 1,000 times as much as the first.

This absurd result of an arbitrary relative contribution of the processes depending on the mass scale used is a consequence of the flaw discussed earlier. For a linear combination of different power laws, it does not follow that an effective exponent is simply a weighted average of the individual exponents<sup>4</sup>.

The cascade model cannot be salvaged by recasting equation (1) in the form  $MR = MR_0 \sum c_i (M/M_0)^{b_i}$ , where the body mass is measured in units of a 'characteristic' body mass,  $M_0$ , and  $MR_0$  is the metabolic rate of an organism of this mass. First, as the power law in question spans 20 orders of magnitude in mass, there is no characteristic mass scale and the choice of value for  $M_0$  is completely arbitrary.

Second, the value of MR calculated for a species of any given body mass will differ with the choice of  $M_0$ , unless the  $c_i$  values are allowed to vary in response (in which case they are neither constants nor are they constrained to sum to 1, as equation (2) requires of control coefficients).

Third, both the relative contributions to MR of each of the  $i$  terms in the summation, and the relative values of MR calculated for species of different masses, depend on the choice of  $M_0$ . For any given set of body masses, the effective slopes of the

log-log mass-metabolic-rate plots, such as those in ref. 1 and based on this new equation, will exhibit any value between the lowest and highest  $b_i$  value, depending on the value that is chosen for  $M_0$ .

**Jayanth R. Banavar\***, **John Damuth†**, **Amos Maritan‡**, **Andrea Rinaldo§**

\*Department of Physics, 104 Davey Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802, USA

e-mail: jayanth@phys.psu.edu

†Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106, USA

‡International School for Advanced Studies, 34014 Trieste, and INFN & Abdus Salam International Center for Theoretical Physics, 34014 Trieste, Italy

§Dipartimento di Ingegneria Idraulica, Marittima e Geotecnica, Università di Padova, 35131 Padova, Italy

1. Darveau, C.-A., Suarez, R. K., Andrews, R. D. & Hochachka, P. W. *Nature* **417**, 166–170 (2002).
2. Weibel, E. R. *Nature* **417**, 131–132 (2002).
3. Burness, G. P. *Science* **296**, 1245–1247 (2002).
4. Hochachka, P. W. & Somero, G. N. *Biochemical Adaptation: Mechanism and Process in Physiological Evolution* (Oxford Univ. Press, Oxford, 2001).

*Darveau et al. reply* — West *et al.* and Banavar *et al.* criticize our results on mathematical grounds, but they overlook the consistency of our multiple-cause model (concept) of metabolic scaling<sup>1</sup> with what is known from biochemical<sup>2</sup> and physiological<sup>3</sup> analysis of metabolic control. Their single-cause explanations<sup>4,5</sup> are based on the assumption that whole-body metabolism in animals is exclusively supply-limited, whereas there are many factors that together explain the observed patterns of metabolic scaling<sup>6,7</sup>. Our concept can accommodate these multiple causes, the range of metabolic scaling exponents observed in various taxa<sup>8</sup>, and variation in exponents due to physiological state<sup>7</sup>.

Allometric equations are mathematical descriptions of empirical relationships, rather than derived physical laws<sup>6</sup>. Our equation<sup>1</sup> is a first approximation that attempts to express our concept in mathematical terms. It does not distinguish between energy-demand processes that occur in parallel and supply processes that operate in series. It suffers from a semantic flaw that imposes units on the control coefficient,  $c_i$ .

In a modified equation (J. Endelman) to determine the basal metabolic rate,  $BMR = MR_0 \sum c_i (M/M_0)^{b_i}$ , where  $MR_0$  is the characteristic metabolic rate of an animal with a characteristic body mass  $M_0$ ,  $c_i$  is rendered dimensionless while the exact meaning of the original equation<sup>1</sup> is retained. With  $M_0$  of 1 unit mass,  $MR_0$  now takes the

place of the value  $a$ , as found in the standard scaling equation<sup>6</sup> and in our original. For mammalian maximum metabolic rate, MMR, the same equation applies with a roughly tenfold higher  $MR_0$ . We were able to find the relevant  $b_i$  values and estimate  $c_i$  for various processes in mammals to demonstrate the utility of our model.

Although using mammalian data precludes extrapolation to non-mammalian species, our concept can be used to understand metabolic scaling in other taxa. Our equation is not a power-law function, but yields meaningful results when a biologically realistic range of  $b_i$  values is used in simulations. The examples we used yield results that are indistinguishable from power functions, reflected in  $r^2$  values that are greater than 0.999. Lower  $r^2$  values result when  $b_i$  values outside the biological range are used.

The inherent limitations of the data, and estimates based on them, offer new directions for experiments, and the shortcomings of our equation highlight the need for better ways to express our multiple-cause model. Branching distributive structures and supply limitations<sup>4,5</sup> may contribute to metabolic scaling, although supply limitations contribute minimally to BMR, which scales with an exponent that is close to 0.75. Supply limitations have a greater influence on MMR<sup>3,9</sup>, but the allometric exponent for this is paradoxically higher<sup>7</sup>. These and the many factors that contribute to the allometric scaling of metabolic rates<sup>6,7,10</sup>, as well as the observation that cellular metabolic rates *in vitro* decline with increasing body mass<sup>10,11</sup>, should give pause to advocates of single-cause explanations for metabolic scaling.

**Charles-A. Darveau\***, **Raul K. Suarez†**, **Russel D. Andrews‡**, **Peter W. Hochachka\***

\*Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

†Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106-9610, USA

e-mail: suarez@lifesci.ucsb.edu

‡Institute of Marine Science, University of Alaska, Fairbanks and the Alaska Sea Life Center, Seward, Alaska 99664, USA

1. Darveau, C.-A., Suarez, R. K., Andrews, R. D. & Hochachka, P. W. *Nature* **417**, 166–170 (2002).
2. Snell, K. (ed.) *Understanding the Control of Metabolism* (Portland, London, 1997).
3. Jones, J. H. *Comp. Biochem. Physiol.* **120 B**, 125–138 (1998).
4. Banavar, J. R., Damuth, J., Maritan, A. & Rinaldo, A. *Proc. Natl Acad. Sci. USA* **99**, 10506–10509 (2002).
5. West, G., Brown, J. & Enquist, B. *Science* **276**, 122–126 (1997).
6. Schmidt-Nielsen, K. *Scaling: Why is Animal Size So Important?* (Cambridge Univ. Press, Cambridge, UK, 1984).
7. Weibel, E. R. *Symmorphosis: On Form and Function in Shaping Life* (Harvard Univ. Press, Cambridge, Massachusetts, 2000).
8. Lovegrove, B. G. *Am. Nat.* **156**, 201–219 (2000).
9. Di Prampero, P. E. *J. Exp. Biol.* **115**, 319–331 (1985).
10. Hulbert, A. J. & Else, P. L. *Annu. Rev. Physiol.* **62**, 207–235 (2000).
11. Porter, R. K. *Cell. Mol. Life Sci.* **58**, 815–822 (2001).

# Dynein structure and power stroke

Stan A. Burgess\*, Matt L. Walker\*, Hitoshi Sakakibara†, Peter J. Knight\* & Kazuhiro Oiwa†

\* Astbury Centre for Structural Molecular Biology & School of Biomedical Sciences, University of Leeds, Leeds, LS2 9JT, UK

† Kansai Advanced Research Centre, Communications Research Laboratory, Kobe, 651-2492, Japan

**Dynein ATPases are microtubule motors that are critical to diverse processes such as vesicle transport and the beating of sperm tails; however, their mechanism of force generation is unknown. Each dynein comprises a head, from which a stalk and a stem emerge. Here we use electron microscopy and image processing to reveal new structural details of dynein c, an isoform from *Chlamydomonas reinhardtii* flagella, at the start and end of its power stroke. Both stem and stalk are flexible, and the stem connects to the head by means of a linker approximately 10 nm long that we propose lies across the head. With both ADP and vanadate bound, the stem and stalk emerge from the head 10 nm apart. However, without nucleotide they emerge much closer together owing to a change in linker orientation, and the coiled-coil stalk becomes stiffer. The net result is a shortening of the molecule coupled to an approximately 15-nm displacement of the tip of the stalk. These changes indicate a mechanism for the dynein power stroke.**

Dyneins fall into two principal classes<sup>1</sup>. Isoforms in the 9 + 2 axoneme produce the propagating bending motions of cilia and flagella<sup>2–5</sup>. Cytoplasmic isoforms drive a variety of fundamental cellular processes, including nuclear migration, organization of the mitotic spindle, chromosome separation during mitosis, and the positioning and function of many intracellular organelles<sup>6</sup>. Each dynein is a complex of between 1 and 3 heavy chains, each with a relative molecular mass greater than 500,000, together with a number of intermediate and light chains<sup>7</sup>. Each heavy chain contains the sites of ATP hydrolysis and microtubule binding, and thus constitutes the fundamental motor unit<sup>8</sup>. Electron microscopy has established that the heavy chain folds to form a globular head with two elongated structures, the stalk and stem, emerging from it<sup>9–11</sup>. These two structures bind the microtubule track and cargo, respectively<sup>7,12,13</sup>. The stalk is up to 15 nm in length in some species<sup>10</sup> and is most probably an anti-parallel coiled-coil<sup>14</sup>. A small globular domain at its tip is the ATP-sensitive microtubule-binding domain<sup>14–16</sup>. The stalk and head are formed by the carboxyterminal two-thirds of the heavy chain<sup>14,17</sup>. The stem, which also binds the intermediate and light chains<sup>7</sup>, is formed by the aminoterminal region and contains numerous short stretches with predicted coiled-coil structure<sup>18,19</sup>.

## Dyneins are AAA proteins

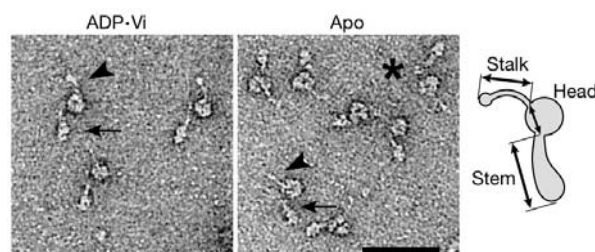
Dynein heavy chains belong to the AAA superfamily of mechano-enzymes (ATPases associated with diverse cellular activities)<sup>20</sup>, each containing six tandemly linked AAA modules within the head<sup>12,13,21</sup>. Electron microscopy studies have indicated that the head is a ring-like arrangement of domains<sup>17</sup>, consistent with the finding that other members of the AAA superfamily form ring-like structures<sup>21</sup>. Only four of the dynein AAA modules (numbers 1–4) have conserved P-loop signatures<sup>18,19</sup>, compatible with functional nucleotide binding at these sites<sup>22</sup>. However, AAA1, nearest the N terminus, is the only one that hydrolyses ATP at an appreciable rate<sup>23</sup>. The function of the remaining three sites is unknown but they are thought to act in a regulatory role. This is most probably mediated by binding ADP, as, for some dyneins, ATP-induced microtubule translocation *in vitro* either requires ADP or is accelerated by it<sup>24,25</sup>.

In the cycle of ATP hydrolysis, release of ADP and phosphate (Pi) from dynein is associated with the power stroke<sup>26</sup>. In the presence of ATP and vanadate (Vi), dynein forms a stable complex (ADP•Vi-dynein), thought to mimic the ADP•Pi state and hence the pre-power-stroke conformation of the motor<sup>26</sup>. Its structure in the absence of nucleotide (apo-dynein) is thought to represent the post-power-stroke conformation<sup>26</sup>. Electron microscope studies of

the outer-arm dynein complex *in situ* have shown that these two biochemical states have distinct conformations<sup>27,28</sup>, and both have been found in motile flagella in the presence of ATP, establishing that they exist within the chemo-mechanical cycle of dynein<sup>28</sup>. Thus, the transition from ADP•Vi states to apo-states is analogous to product release and the power stroke. However, each outer-arm complex comprises multiple dynein heavy chains that were not resolved in these structural studies, so the conformational change in an individual heavy chain remains unknown. The existing evidence suggests that the mechanism of dynein is likely to be different from those of the other, much smaller, linear motors kinesin and myosin<sup>29</sup>. However, it remains unknown which part of the molecule amplifies the conformational change within the active site in the head into large-scale movement, and how the coiled-coil stalk conveys conformational changes to the ATP-sensitive microtubule-binding domain at the tip of the stalk.

## Dynein c structure at the start and end of the power stroke

Figure 1 shows isolated molecules of a monomeric, flagellar, inner-arm dynein (subspecies c) imaged in both biochemical states using negative stain. These images show that the ADP•Vi molecule has the same general form as the apo-molecule<sup>10,30</sup>, but the latter is more compact. We have used single-particle image processing<sup>31</sup> to reveal this change more clearly. Dynein c adopts two characteristic orientations on the carbon support film, providing two views (left and right) related by a roughly 180° rotation about the long axis of the molecule (Fig. 2a–h). Flexibility within the molecule produces a



**Figure 1** Negatively stained molecules of ADP•Vi-dynein c (left panel) and apo-dynein c (middle panel) from *C. reinhardtii* flagella. The preparation is monodisperse. Stems (arrows) and stalks (arrowheads) are indicated. The asterisk shows a molecule with a longer stem (see Fig. 2i). Scale bar, 50 nm. The right panel is a diagram of dynein c defining the stem and stalk, and the separation between their emergence points from the head.



continuous range of conformations of the stem and stalk (see Supplementary movies 1–4; see also <http://www.leeds.ac.uk/bms/research/muscle/dynein>) causing most of the relatively smaller stalk to become smeared out in these averages (Fig. 2a, b, e, f). In both biochemical states the stem most commonly emerges from the left side of the head (Fig. 2a, e). At its base the stem is wide, narrowing to a flexible neck about 2 nm wide and about 8 nm long, which bends sharply very close to the head. A second alignment, using just the part of each image containing the head, reveals head details optimally (Fig. 2c, d, g, h). These show that left views of the head are not mirror images of right views in either biochemical state, indicating that the two faces of the head are different.

Comparisons between ADP•Vi- and apo-molecules reveal a pronounced conformational change within the head region. In left views the ADP•Vi head is roughly circular (Fig. 2c), with a diameter of approximately 13 nm, whereas the apo-head has a markedly different shape and a prominent central stain deposit (Fig. 2g). Small variations in stain accumulation or viewing angle produce subtle differences in appearance between individual heads, which limit the detail in these global averages. Further classification of left views therefore reveals additional detail of the conformational change (Fig. 3a, b). A central stain deposit occasionally visible in ADP•Vi heads (Fig. 3a) has greater prominence in apo-heads (Fig. 3b). A similar difference is also apparent in right views (Fig. 2d, h). These observations suggest an opening or unblocking of a stain-filled channel. Right views also show three distinct domains arranged around one edge of the apo-head (asterisks), not seen in the ADP•Vi head, indicating that changes also occur around the periphery.

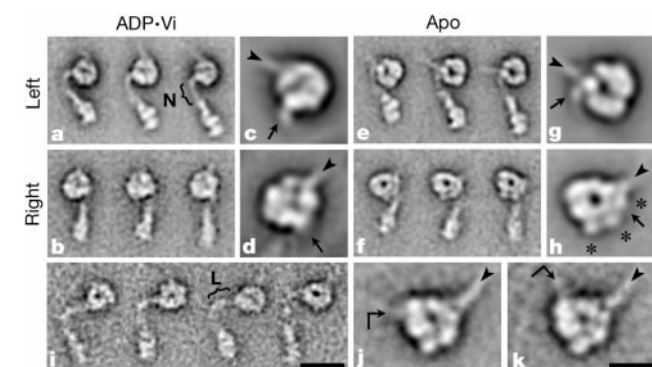
### Movement of the stem

The most marked change in the molecule between different states is that stem and stalk are closer together in apo-molecules than in ADP•Vi molecules (compare Fig. 2g, h with c, d). The separation between their emergence points from the head in left views changes by 6.3 nm (mean chord length: ADP•Vi =  $9.9 \pm 1.3$  nm (s.d.),  $n = 1,235$ ; apo =  $3.6 \pm 0.9$  nm,  $n = 1,221$ ; measured as in Fig. 1 diagram). As there is little change in length of either the stalk

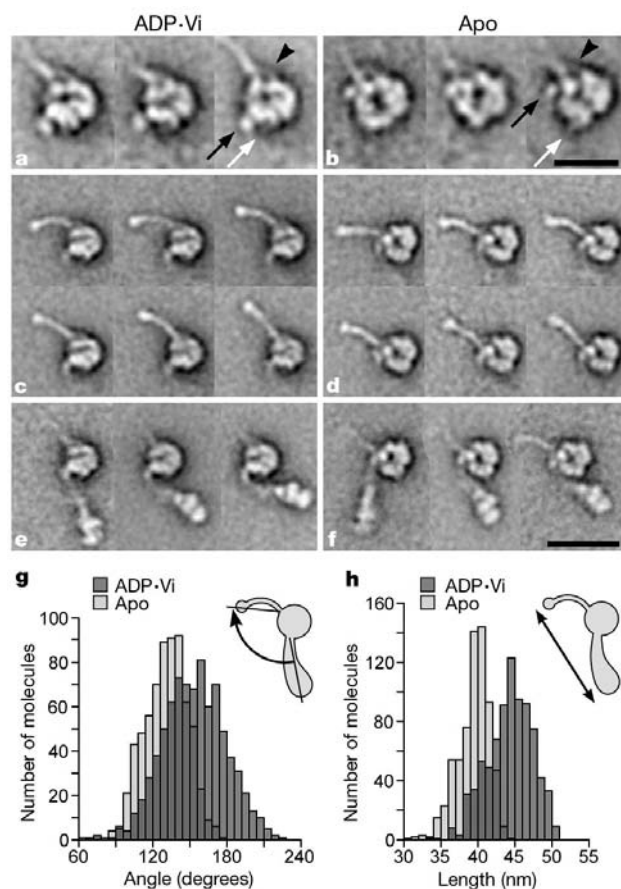
(mean chord length: ADP•Vi =  $15.4 \pm 0.9$  nm,  $n = 908$ ; apo =  $15.5 \pm 1.0$  nm,  $n = 721$ ) or the stem (ADP•Vi =  $26.4 \pm 1.2$  nm,  $n = 1,351$ ; apo =  $25.0 \pm 1.0$  nm,  $n = 1,351$ ; see Fig. 1 diagram) the origin of this change must be a translational movement between these two domains. This also brings the neck and head into closer proximity (for example, compare Fig. 3e with Fig. 3f) but has little effect on the distribution of stem chord angles (standard deviation in ADP•Vi =  $18^\circ$ ,  $n = 1,351$ ; apo =  $16^\circ$ ,  $n = 1,351$ ). A fine, tapering structure on the head (Fig. 3a, b, white arrows) provides a landmark, relative to which the emergence point of the stem clearly changes (Fig. 3a, b, black arrows), but the stalk does not. Thus the origin of the movement is a displacement of the stem relative to the head and stalk.

### Stem and head are connected by a linker domain

Clues to the mechanism producing these changes arise from an analysis of rare, longer-stemmed dynein c molecules (Fig. 1, asterisk). Analysis revealed no consistent differences between ADP•Vi- and apo-molecules in this small data set, so we combined the images to improve the signal to noise ratio. Despite their structural heterogeneity, averages reveal several important features of the organization of the molecule (Fig. 2i–k). There is often a sharp, approximately  $90^\circ$  bend in the stem, at a position corre-



**Figure 2** Characteristic views of dynein c revealed by single-particle analysis. Images are selected class averages of the most common views. **a–d**, ADP•Vi-dynein in left views (**a**) and right views (**b**); after head alignment in left view (**c**) and right view (**d**). **e–h**, Apo-dynein c in left views (**e**) and right views (**f**); after head alignment in left view (**g**) and right view (**h**). **i–k**, Selected class averages of rare molecules showing the linker (L) (**i**) and head substructure after head alignment (**j**, **k**). The location of stalk (arrowheads) and apparent stem attachment (arrows) on the head are indicated. N indicates the neck; asterisks indicate three subdomains. Left views represent about 50% of ADP•Vi- and apo-molecules; right views about 12%. Rare molecules represent about 6% of ADP•Vi- and apo-molecules. The numbers of images in each class average in **a**, **b**, **e** and **f** vary between 20 and 43, and in **i** between 5 and 6. The number of images in averages after head alignment are: **c**, 1,733; **d**, 361; **g**, 1,658; **h**, 346; **j**, 38 (18 ADP•Vi + 20 apo); and **k**, 23 (12 ADP•Vi + 11 apo). Scale bars: **a**, **b**, **e**, **f**, **i**, 25 nm; **c**, **d**, **g**, **h**, **j**, **k**, 10 nm.



**Figure 3** Substructure and flexibility within left views of dynein c. Head-aligned molecules (that is, those contributing to Fig. 2c, g) were further classified independently according to their heads, stalks and stems, using masks to isolate these structures. **a–f**, Image averages of selected classes of: ADP•Vi heads (**a**), apo-heads (**b**), ADP•Vi stalks (**c**), apo-stalks (**d**), ADP•Vi stems (**e**) and apo-stems (**f**). **a**, **b**, The stalk bifurcates near its point of attachment to the head (arrowheads). A fine structure emerges from the head (white arrows). The emergence point of the stem is indicated (black arrows). The number of images in class averages vary between 14 and 44. **g**, Distribution of stem–stalk angles. **h**, Distribution of molecular lengths. Scale bars: **a**, **b**, 15 nm; **c–f**, 25 nm.

sponding to the bend seen in the neck of typical molecules. Between this bend and the head is a hitherto unseen structure, similar in width to the neck and about 10 nm long, which we call the linker (Fig. 2i). Thus the stem consists of three distinct domains: base, neck and linker. The similarity of the head (Fig. 2j, k) to right views of typical apo-heads (Fig. 2h) shows that the head adopts the same orientation. However, the linker emerges from the left side of the head, diametrically opposite where the neck typically emerges (compare Fig. 2j, k to h; see also Supplementary movies 5 and 6). The straightforward interpretation is that the linker is normally docked onto one face of the head but in these rare molecules it is undocked, revealing the true site of insertion of the stem into the head. The docked linker lies across the upper face in right views, as the lower face, which is embedded in stain, appears unaffected by dissociation of the linker. Consistent with this, the neck is poorly defined in right views of typical molecules, indicating that it lies sufficiently far above the carbon substrate that it is above the stain. Thus movement of the docked stem between the ADP•Vi- and apo-conformations implies a swinging movement of the linker across the head.

### Conformational changes in the stalk

The transition between ADP•Vi states and apo-states also causes structural changes in the stalk. This structure is revealed optimally by image classification of left views after head alignment to bring the bases of the stalks into register (Fig. 3c, d). The stalk is uniformly narrow (about 2 nm), except for a distinct enlargement at its tip. Where it joins the head (Fig. 3a, b, arrowheads), the stalk is resolved into two separate strands. These observations are compatible with other evidence that the stalk consists of two  $\alpha$ -helices interacting as an anti-parallel coiled-coil, with the ATP-sensitive microtubule-binding domain at the tip<sup>11,14–16,32</sup>. Flexibility of the stalk produces a continuous range of conformations in both biochemical states, but the distribution of stalk chord angles indicates that the extent of flexibility is altered by loss of products. The standard deviation of the distribution (measured relative to the head) changes from 20° in ADP•Vi molecules ( $n = 908$ ) to 11° in apo-molecules ( $n = 721$ ),

indicating a stiffening of the stalk in apo-molecules. This is due, at least in part, to a structural change within the stalk, as ADP•Vi stalks are gently curved when flexed (that is, pointing leftwards; Fig. 3c, upper row) and straight when extended (pointing upwards; Fig. 3c, lower row), whereas apo-stalks (Fig. 3d) are always straight except for a slight kink about 5 nm from the tip (see Supplementary movies 1 and 2).

### Power stroke of dynein

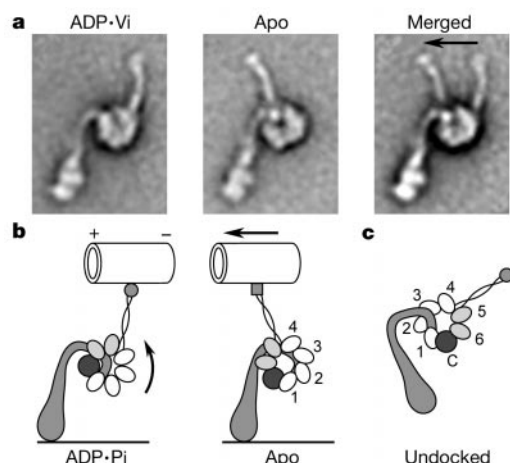
The conformational change seen in left views alters the angle between stem and stalk, and changes the overall length of the molecule. The mean angle between stem and stalk (Fig. 3g) decreases from 160° (ADP•Vi) to 136° (apo). The decrease in standard deviation of this distribution, from 25° (ADP•Vi) to 17° (apo) is accounted for by the stiffening of the stalk described above. There is no correlation between the angles of stem and stalk in individual molecules (data not shown), indicating that their movements are not coupled. The mean length of dynein c decreases markedly from 45.3 to 40.4 nm (Fig. 3h), producing the more compact apo-molecule apparent in raw images.

To examine the magnitude of the power stroke we compared images of left views showing the mean conformation of stem and stalk in ADP•Vi- and apo-molecules (Fig. 4a). Alignment of their stems suggests a mean displacement of 15 nm of the tip of the stalk (see Supplementary movie 7). A change in head orientation may also contribute to the apparent (projected) size of this displacement. This power stroke is longer than the 8-nm steps seen under load in optical trap experiments<sup>30</sup> that reflect the spacing of successive binding sites on the microtubule. Moreover, the flexibility of both neck and stalk implies that power stroke amplitudes can cover a wide range around this mean.

### Discussion

How might these new structural data be reconciled with earlier findings? The appearance of the head of dynein c supports the idea of a toroidal arrangement of domains suggested previously<sup>12,13,21,33</sup>. The head is also large enough to accommodate at least six AAA domains arranged in a planar ring, as described by a hypothetical atomic model of dynein<sup>33</sup>. However, this model predicts a similar appearance of the two faces and six-fold rotational symmetry—neither is seen in our images. A possible origin of these discrepancies is that a seventh AAA module<sup>34</sup> and the C-terminal sequence<sup>12</sup> may contribute additional domains within the head. Furthermore, by lying across one face of the head, the linker would make the two faces appear different. Undocking of the linker from the head may explain two previous observations. First, images showing that the stem but not the stalk elongates when additional tension is applied to the multimeric outer-arm complex<sup>9</sup> can now be interpreted as domain movements within individual dynein heavy chains. Second, isolated molecules of other dynein species<sup>9–11</sup> display a structural heterogeneity and unexpected length similar to the rare undocked molecules of dynein c, suggesting that the linker has a weaker association in these species.

It has been proposed that stem, stalk or both act as mechanical levers, amplifying conformational changes originating within the head<sup>12,13,15,21,32</sup>. Our data show that amplification is achieved primarily through a change in head–stem orientation. This refines the conclusion from earlier studies on multimeric flagellar outer-arm dynein complexes *in situ*<sup>11,28</sup>. Flexibility in the stalk and neck suggest that these may be important compliant, although inextensible, linkages that are capable of storing elastic strain energy when the molecule develops force against a load. This ability may underlie the ATP-induced propagation of a zone of activity along isolated groups of flagellar microtubules<sup>3</sup>. Stiffening of the apo-stalk is compatible with the increased stiffness of nucleotide-free flagella over those bathed in ATP and vanadate<sup>35</sup>. The affinity of the microtubule-binding domain for its track is higher in apo-dynein than in



**Figure 4** Structure and power stroke of dynein c. **a**, Left views of dynein c showing the mean conformation of the stem and stalk relative to the head. Images were made by compositing class averages having the mean stem and stalk angles. Images are shown after alignment of their stems. The arrow indicates 15 nm. **b**, Speculative structure of the molecule suggesting an origin of the power stroke and how this might be converted into microtubule movement. The six AAA modules (1–6) and C-terminal sequence (C) are indicated. The microtubule track (not drawn to scale) is illustrated as moving under zero load with its plus end leading<sup>30</sup>. **c**, Molecule showing the stem undocked from the head, shown as a right view for comparison with Fig. 2i–k. The same stem structure has been used in all illustrations.

ADP•Vi-dynein<sup>36</sup>, and microtubules activate the ATPase<sup>37</sup>, implying two-way communication between the tip of the stalk and the ATPase site in the head<sup>14,32</sup>. Our finding that the conformation of the stalk itself changes indicates that this mechanism involves modifications in the structure of the coiled-coil.

We suggest a new model for the structure and mechanism of dynein c (Fig. 4b). Six AAA domains and a C-terminal domain form a planar ring<sup>12</sup>, with the stalk emerging between AAA4 and -5, and a hook-shaped stem joining AAA1. The rare undocked appearance (Fig. 2i–k) can be understood as arising from a change in structure at the junction between the linker and AAA1, accompanied by undocking of the linker from the head. In molecules adsorbed to the carbon in the right view this allows the entire stem to flip over (Fig. 4c). The possible physiological relevance of such molecules is currently unknown. In the typical ADP•Vi molecule (Fig. 4b, left), the docked linker occludes the central channel and the stem emerges from the head far from the stalk. Microtubule movement is initiated by tight binding to the tip of the stalk, which promotes a concerted conformational change in AAA1–4, thereby activating release of ADP and phosphate from AAA1, the probable site of ATP hydrolysis<sup>33</sup>. Rigid coupling between AAA1 and linker causes a rolling of the head towards the stem, which translates the microtubule by 15 nm under zero load conditions. The resulting movement of AAA2–4 towards the linker increases the prominence of the central channel (Fig. 4b, right). Movement between AAA4 and -5 increases stalk stiffness by changing the structure of its junction with them. The approach of the apo-linker to this region may also be involved. Presumably, ATP binding to AAA1 reduces the microtubule-binding affinity without a complete reversal of this sequence of events. In support of the mechanism, both ATP and ADP binding within AAA1–4 (ref. 22) is required for optimal dynein function<sup>24</sup>. Nucleotide in these sites might favour conformational switching on the communication pathway. This mechanism offers the intriguing possibility that functionally significant interactions between AAA2–4 and the linker may occur, and be modulated by nucleotide. Our results identify that a change in head–linker orientation is fundamental in the power stroke of dynein. □

## Methods

Dynein c was purified from *C. reinhardtii* as described<sup>30</sup>, except that to improve the quality of negative staining, the elution buffer for the final anion exchange column was a KCl gradient in 30 mM MOPS, 5 mM MgCl<sub>2</sub>, 1 mM EGTA and 0.1 mM dithiothreitol, pH 7.4 (MMED buffer). Stock dynein samples (approximately 0.65 μM protein and about 200 mM KCl) were diluted 40-fold into MMED buffer containing 200 mM KCl. To generate ADP•Vi molecules, 100 μM ATP and 50 μM sodium orthovanadate were added at 4 °C for 30 min<sup>38</sup>. The specimen was applied to carbon grids treated with ultraviolet light, as described<sup>39</sup>. Before staining with 1% uranyl acetate<sup>39</sup>, grids were rinsed with MMED buffer without KCl to improve staining. Micrographs were taken at a nominal magnification of ×40,000 and calibrated using the paramyosin spacing of 14.4 nm<sup>40</sup>. A total of 4,900 ADP•Vi molecules and 3,057 apo-molecules were selected interactively and subjected to single-particle image processing using procedures written in the SPIDER suite of programs<sup>31</sup>. Images were aligned by reference-free algorithms and classified into homogeneous groups by K-means clustering, as described<sup>41</sup>. Length and angle measurements were made from class averages and weighted according to the number of constituent images before calculating mean and standard deviation. *n* values are given as the total number of images, rather than classes.

Received 6 September; accepted 17 December 2002; doi:10.1038/nature01377.

- Gibbons, I. R. Dynein family of motor proteins: present status and future questions. *Cell Motil. Cytoskel.* **32**, 136–144 (1995).
- Gibbons, I. R. & Rowe, A. Dynein: a protein with adenosine triphosphatase activity from cilia. *Science* **149**, 424–426 (1965).
- Vernon, G. G. & Woolley, D. M. The propagation of a zone of activation along groups of flagellar doublet microtubules. *Exp. Cell Res.* **220**, 482–494 (1995).
- Woolley, D. M. The molecular motors of cilia and flagella. *Essays Biochem.* **35**, 103–115 (2000).
- DiBella, L. M. & King, S. M. Dynein motors of the *Chlamydomonas* flagellum. *Int. Rev. Cytol.* **210**, 227–268 (2001).
- Karki, S. & Holzbaur, L. F. Cytoplasmic dynein and dynactin in cell division and intracellular transport. *Curr. Opin. Cell Biol.* **11**, 45–53 (1999).
- King, S. M. The dynein microtubule motor. *Biochim. Biophys. Acta* **1496**, 60–75 (2000).

- Koonce, M. P. & Samso, M. Overexpression of cytoplasmic dynein's globular head causes a collapse of the interphase microtubule network in *Dictyostelium*. *Mol. Biol. Cell* **7**, 935–948 (1996).
- Goodenough, U. W. & Heuser, J. Structural comparison of purified dynein proteins with *in situ* dynein arms. *J. Mol. Biol.* **180**, 1083–1118 (1984).
- Goodenough, U. W. *et al.* High-pressure liquid chromatography fractionation of *Chlamydomonas* dynein extracts and characterization of inner-arm dynein subunits. *J. Mol. Biol.* **194**, 481–494 (1987).
- Goodenough, U. W. & Heuser, J. E. in *Cell Movement: the Dynein ATPases* Vol. 1 (eds Warner, F. D., Satir, P. & Gibbons, I. R.) 121–140 (Alan Liss, New York, 1989).
- King, S. M. AAA domains and organization of the dynein motor unit. *J. Cell Sci.* **113**, 2521–2526 (2000).
- Asai, D. J. & Koonce, M. P. The dynein heavy chain: structure, mechanics and evolution. *Trends Cell Biol.* **11**, 196–202 (2001).
- Gee, M. A., Heuser, J. E. & Vallee, R. B. An extended microtubule-binding structure within the dynein motor domain. *Nature* **390**, 636–639 (1997).
- Vallee, R. B. & Gee, M. A. Make room for dynein. *Trends Cell Biol.* **8**, 490–494 (1998).
- Koonce, M. P. & Tikhonenko, I. Functional elements within the dynein microtubule-binding domain. *Mol. Biol. Cell* **11**, 523–529 (2000).
- Samsó, M., Radermacher, M., Frank, J. & Koonce, M. P. Structural characterization of a dynein motor domain. *J. Mol. Biol.* **276**, 927–937 (1998).
- Gibbons, I. R., Gibbons, B. H., Mocz, G. & Asai, D. Multiple nucleotide-binding sites in the sequence of dynein β heavy chain. *Nature* **352**, 640–643 (1991).
- Ogawa, K. Four ATP-binding sites in the midregion of the β heavy chain of dynein. *Nature* **352**, 643–645 (1991).
- Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA + : a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
- Vale, R. D. AAA proteins: lords of the ring. *J. Cell Biol.* **150**, F13–F19 (2000).
- Mocz, G. & Gibbons, I. R. Phase partition analysis of nucleotide binding to axonemal dynein. *Biochemistry* **35**, 9204–9211 (1996).
- Gibbons, I. R. *et al.* Photosensitized cleavage of dynein heavy chains. *J. Biol. Chem.* **262**, 2780–2786 (1987).
- Yagi, T. ADP-dependent microtubule translocation by flagellar inner-arm dyneins. *Cell Struct. Funct.* **25**, 263–267 (2000).
- Shiroguchi, K. & Toyoshima, Y. Regulation of monomeric dynein activity by ATP and ADP concentrations. *Cell Motil. Cytoskel.* **49**, 189–199 (2001).
- Johnson, K. A. Pathway of the microtubule-dynein ATPase and the structure of dynein: a comparison with actomyosin. *Annu. Rev. Biophys. Biophys. Chem.* **14**, 161–188 (1985).
- Goodenough, U. W. & Heuser, J. E. Substructure of the outer dynein arm. *J. Cell Biol.* **95**, 798–815 (1982).
- Burgess, S. A. Rigor and relaxed outer dynein arms in replicas of cryofixed motile flagella. *J. Mol. Biol.* **250**, 52–63 (1995).
- Vale, R. D. & Milligan, R. A. The way things move: looking under the hood at molecular motor proteins. *Science* **288**, 88–95 (2000).
- Sakakibara, H., Kojima, H., Sakai, Y., Katayama, E. & Oiwa, K. Inner-arm dynein c of *Chlamydomonas* flagella is a single-headed processive motor. *Nature* **400**, 586–590 (1999).
- Frank, J. *Three-dimensional Electron Microscopy of Macromolecular Assemblies* (Academic, New York, 1996).
- Gee, M. A. & Vallee, R. B. The role of the dynein stalk in cytoplasmic and flagellar motility. *Eur. Biophys. J.* **27**, 466–473 (1998).
- Mocz, G. & Gibbons, I. R. Model of the motor component of dynein heavy chain based on homology to the AAA family of oligomeric ATPases. *Structure* **9**, 93–103 (2001).
- Fan, J. & Amos, L. A. Antibodies to cytoplasmic dynein heavy chain map the surface and inhibit motility. *J. Mol. Biol.* **307**, 1317–1327 (2001).
- Okuno, M. Inhibition and relaxation of sea urchin sperm flagella. *J. Cell Biol.* **85**, 712–725 (1980).
- Vale, R. D., Soll, D. R. & Gibbons, I. R. One-dimensional diffusion of microtubules bound to flagellar dynein. *Cell* **59**, 915–925 (1989).
- Omoto, C. K. & Johnson, K. Activation of the dynein adenosine triphosphatase by microtubules. *Biochemistry* **25**, 419–427 (1986).
- Gibbons, I. R. & Mocz, G. Photocatalytic cleavage of proteins with vanadate and other transition metal complexes. *Methods Enzymol.* **196**, 428–442 (1991).
- Walker, M. L. *et al.* Two-headed binding of a processive myosin to F-actin. *Nature* **405**, 804–807 (2000).
- Elliot, A., Offer, G. & Burridge, K. Electron microscopy of myosin molecules from muscle and non-muscle sources. *Proc. R. Soc. Lond. B* **193**, 45–53 (1976).
- Burgess, S. A., Walker, M. L., White, H. D. & Trinick, J. Flexibility within myosin heads revealed by negative stain and single-particle analysis. *J. Cell Biol.* **139**, 675–681 (1997).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com/nature>).

**Acknowledgements** We thank J. Trinick for comments on an earlier draft. This work was supported in part by a NIH grant to J. Trinick and H. White, and by the BBSRC.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to S.A.B. (e-mail: s.a.burgess@leeds.ac.uk).



# The far-ultraviolet signature of the 'missing' baryons in the Local Group of galaxies

Fabrizio Nicastro\*, Andreas Zezas\*, Martin Elvis\*, Smita Mathur†, Fabrizio Fiore‡, Cesare Cecchi-Pestellini\*, Douglas Burke\*, Jeremy Drake\* & Piergiorgio Casella\*

\* Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

† Astronomy Department, The Ohio State University, Columbus, Ohio 43210, USA

‡ Osservatorio Astronomico di Monteporzio, Via Osservatorio Via Frascati 33, Monteporzio-Catone (RM), I-00040 Italy

The number of baryons detected in the low-redshift ( $z < 1$ ) Universe is far smaller than the number detected in corresponding volumes at higher redshifts. Simulations<sup>1–3</sup> of the formation of structure in the Universe show that up to two-thirds of the 'missing' baryons may have escaped detection because of their high temperature and low density. One of the few ways to detect this matter directly is to look for its signature in the form of ultraviolet absorption lines in the spectra of background sources such as quasars. Here we show that the amplitude of the average velocity vector of 'high velocity' O VI ( $O^{5+}$ ) absorption clouds detected in a survey<sup>4</sup> of ultraviolet emission from active galactic nuclei decreases significantly when the vector is transformed to the frames of the Galactic Standard of Rest and the Local Group of galaxies. At least 82 per cent of these absorbers are not associated with any 'high velocity' atomic hydrogen complex in our Galaxy, and are therefore likely to result from a primordial warm-hot intergalactic medium pervading an extended corona around the Milky Way or the Local Group. The total mass of baryons in this medium is estimated to be up to  $\sim 10^{12}$  solar masses, which is of the order of the mass required<sup>5</sup> to dynamically stabilize the Local Group.

The first evidence for an absorption system due to the warm-hot intergalactic medium has been discovered in the X-ray and ultraviolet spectra of the blazar PKS2155–304 (ref. 6), obtained with the Chandra low-energy transmission grating (LETG)<sup>7</sup>, and the Far-Ultraviolet Spectroscopic Explorer (FUSE)<sup>8</sup>. But one line of sight is inadequate to fully define this absorbing medium. FUSE observations of active galactic nuclei show two types of ubiquitous  $z \approx 0$  O VI absorbers, similar to those observed in PKS2155–304: (1) low-velocity O VI clouds (LV-O VI, velocity with respect to the Local Standard of Rest  $|v_{LSR}| < 100 \text{ km s}^{-1}$ )<sup>9–13</sup>; and (2) high-velocity O VI clouds (HV-O VI,  $|v_{LSR}| > 100 \text{ km s}^{-1}$ )<sup>4,14–16</sup>. To investigate the nature of these absorbers we have examined (see Methods section) a sample of active galactic nuclei using publicly available FUSE data. This has allowed us to clearly identify the HV-O VI component with diffuse gas in an extended Galactic corona or in the Local Group.

The O VI velocity distribution in the LSR (Fig. 1, dashed histogram) shows a narrow peak between  $\pm 100 \text{ km s}^{-1}$  (LV-O VI) with a much broader, roughly symmetric distribution (Fig. 1, solid histogram) extending to  $\pm 550 \text{ km s}^{-1}$  (HV-O VI). The bimodality of this distribution suggests that LV- and HV-O VI systems belong to two different populations of absorbers, as previously pointed out in refs 9 and 14. Here we concentrate on the HV-O VI absorbers, and use the LV-O VI absorbers only as a comparison sample. A complete discussion of both LV and HV systems is deferred to a forthcoming paper (F.N., A.Z., M.E. & P. Casella, manuscript in preparation).

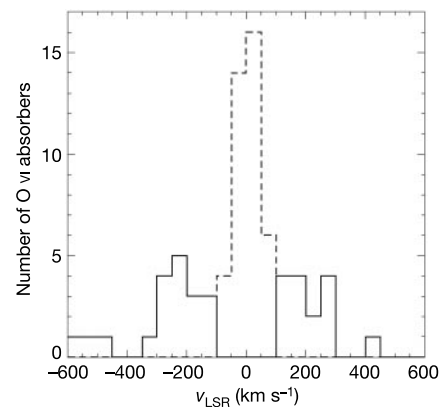
A plot in Galactic coordinates (latitude,  $b$ ; longitude,  $l$ ) of the LSR velocity distributions for the HV (Fig. 2a) absorption systems shows that the hemisphere with  $0^\circ \leq l \leq 180^\circ$  contains only HV-O VI lines

with negative velocities, while the other hemisphere contains mostly HV-O VI absorbers with positive velocity. There are three exceptions: two of these have LSR velocities very close to the threshold velocity of  $|v_{LSR}| = 100 \text{ km s}^{-1}$ , and so may well belong to the LV-O VI population. The third lies at very high latitude, where the concept of longitude becomes meaningless. For comparison, the population of LV-O VI absorbers show no velocity segregation in the LSR.

The red circles in Fig. 2 mark the six HV-O VI absorbers ( $\sim 18\%$  of the sample) which are tentatively identified with 21-cm high-velocity H I clouds or complexes, either because of spatial and velocity coincidence (within broad ranges) with entries in the catalogues of refs 17 and 18, or because they are identified as such in the HV-O VI compilation of ref. 4 (in which  $\sim 20\%$  of HV-O VI are tentatively identified with HV-H I). We stress here that these are the only six HV-O VI absorbers, in either compilation (ours and ref. 4), that may be identified with HV-H I clouds with a known distance, either in our Galaxy's inner halo (with distances less than  $\sim 10 \text{ kpc}$ ), or between us and the two Magellanic Clouds. To be conservative we do not consider further these six HV-O VI absorbers.

The systematic LSR velocity distribution of HV-O VI is consistent with matter that is (1) counter-rotating, with respect to the Galaxy disk rotation, on orbits external to the Sun's orbit, (2) at rest in the Galactic halo, or (3) at rest in the intergalactic space surrounding the Galaxy. The range of radial LSR velocities of the HV-O VI ( $100 < |v_{LSR}|^{HV} < 550 \text{ km s}^{-1}$ ) greatly exceeds the range of observed radial velocities in the Galactic disk or halo, suggesting that the Galaxy-related options (1) and (2) are unlikely. Moreover, some of the HV-O VI velocities exceed a plausible measure of the escape velocity from the Milky Way<sup>19</sup>. Finally, clouds in the Galaxy's halo would probably be rotating on random orbits around the Galaxy's centre, as globular clusters do. The peculiar velocities of these clouds along these orbits would tend to randomize the apparent symmetry induced in the LSR by the circular motion of the Sun in the Galaxy for matter effectively at rest in the halo, as observed in globular clusters. An intergalactic origin, with distances larger than  $\sim 100 \text{ kpc}$ , is more consistent with the data.

The LSR is not a rest frame system for the HV-O VI absorbers (Fig. 2a, Table 1). This is confirmed by translations of the velocity distribution to other convenient rest frames: the Galactic Standard



**Figure 1** The velocity range of HV-O VI absorbers greatly exceeds typical rotational velocities in the Galaxy. Histogram of the HV-O VI (solid line) and LV-O VI (dashed line) velocity distributions in the Local Standard of Rest (LSR). Among the 54 objects of our sample, 45 (83%) show at least one clear O VI absorption component at  $z \approx 0$  at our detection thresholds (7 out of the remaining 9 objects have poor-quality FUSE spectra, with detection thresholds of equivalent width  $\geq 200 \text{ mÅ}$ ). Four show multiple LV and HV absorption. 38 lines of sight show LV-O VI absorption (70% of the sample), and 32 (59% of the sample) show HV-O VI components, with 22 objects showing both. Only four lines of sight show multiple LV or HV absorption.

of Rest (GSR) and the Local Group Standard of Rest (LGSR). The symmetry present in the LSR velocity distribution of the HV-O VI systems (Fig. 2a) disappears in the LGSR (Fig. 2b) and instead appears random, suggesting that the LGSR is the rest frame for the HV-O VI absorbers. The amplitude of the HV-O VI average velocity vector decreases monotonically from the LSR to the GSR, to the LGSR, at which point it is only  $\langle |v_{\text{LGSR}}|^{\text{HV}} \rangle = 32 \text{ km s}^{-1}$  (Table 1), much smaller than the corresponding velocities in the LSR, and close to the FUSE resolution ( $\sim 20 \text{ km s}^{-1}$  at  $1,032 \text{ \AA}$ ). At the same time, the direction of the average vector becomes very poorly constrained (Table 1). Both results suggest that GSR and LGSR are preferred rest frames for the HV-O VI absorbers. This locates the population of HV absorbers in intergalactic space, possibly associated with an extended Galactic corona or Local Group medium.

We checked the validity of this method against our control sample of LV-O VI absorbers, for which a nearby Galactic origin has been claimed<sup>9–13</sup>. The velocity signs of LV-O VI are distributed randomly in the LSR, but show a high degree of segregation in the GSR, with positive velocities segregated in the  $0^\circ \leq l \leq 180^\circ$  hemisphere, as expected for clouds co-rotating with the Galactic disk. Furthermore, the amplitude of the average velocity vector of the LV-O VI increases monotonically from the LSR to the LGSR, from  $5 \text{ km s}^{-1}$  up to  $71 \text{ km s}^{-1}$ , opposite to what is observed for the HV-O VI absorbers and again confirming the Galactic origin of the LV-O VI absorbers.

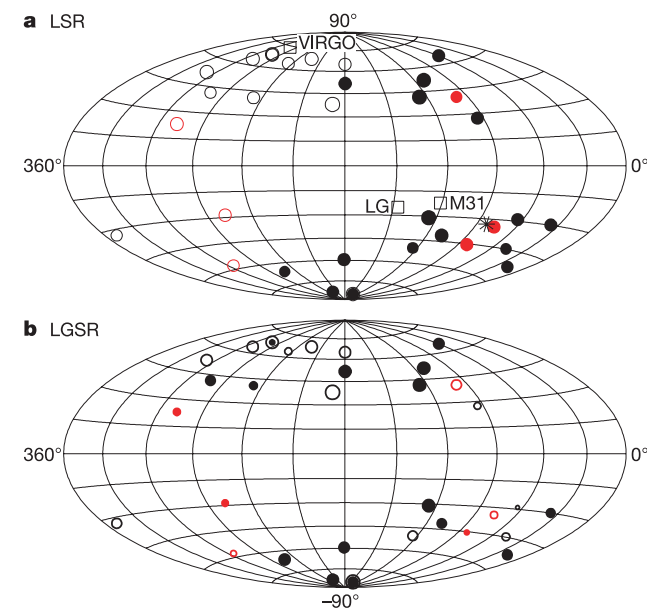
There are several pieces of evidence, both observational and theoretical, supporting the identification of the HV systems with tenuous and diffuse warm-hot gas filling our Local Group. Recent ‘constrained’ hydrodynamical simulations for the formation of structures in our own supercluster environment<sup>20,21</sup> predict a major reservoir of baryons in the local environment in the form of filaments of the warm-hot intergalactic medium. The Local Group should be embedded in one of these filaments. The net motion of our Galaxy toward M31, in the ‘local filament’ (on average at rest in the LGSR), would produce an apparent bulk

motion of the filament, towards our Galaxy, in the direction of M31. This is similar to what is observed for the velocity field HV-O VI absorbers. The inverse of the mean GSR velocity vector of the HV-O VI absorbers (Table 1) has  $\langle |v_{\text{GSR}}|^{\text{HV}} \rangle = -64 \text{ km s}^{-1}$ , and Galactic coordinates of  $l = 76^\circ$ ,  $b = -38^\circ$ , fairly close to M31:  $l = 121^\circ$ ,  $b = -22^\circ$ .

Estimates of the dynamical mass of our Local Group exceed the measured mass of its visible constituents by more than a factor of two (see, for example, ref. 5). A total mass of  $\sim 10^{12}$  solar masses would be needed to stabilize, dynamically, the Local Group. From the combined analysis of the absorption features from highly ionized gas in the ultraviolet (HV-O VI) and X-ray (O VII, O VIII and Ne IX) spectra of the blazar PKS2155–304, it has been<sup>6</sup> concluded that this absorber has to fill the intergalactic space surrounding the Galaxy. This system has an electron density of the order of  $n_e \approx (4\text{--}6) \times 10^6 \text{ cm}^{-3}$ , an equivalent hydrogen column density of  $N_H \approx 4.5 \times 10^{19} ([\text{O}/\text{H}]_{0.3})^{-1} \text{ cm}^{-2}$  (where  $[\text{O}/\text{H}]_{0.3}$  is the O/H metallicity ratio in units of 0.3 times the solar value of  $7.4 \times 10^{-4}$ )<sup>22</sup> and so a linear size along the line of sight of  $D \approx (2\text{--}4) \times ([\text{O}/\text{H}]_{0.3})^{-1} \text{ Mpc}$ . If this particular line of sight is representative of the entire Local Group warm-hot intergalactic medium, then assuming a transverse size of  $1 \times ([\text{O}/\text{H}]_{0.3})^{-1} \text{ Mpc}$  gives a total baryonic mass of  $(0.6\text{--}2) \times ([\text{O}/\text{H}]_{0.3})^{-3} \times 10^{12}$  solar masses, whose upper boundary would be sufficient to stabilize the Local Group.

Additional evidence supporting the Local Group origin of the HV-O VI comes from the velocity distribution of the isolated compact high-velocity clouds of H I (CHVCs), studied in the 21-cm H I emission line<sup>23–26</sup>. On the basis of statistical and theoretical arguments similar to those used here for the HV-O VI absorbers, CHVCs have been proposed to originate at large ( $d > 100 \text{ kpc}$ ) characteristic distances<sup>23,24,27</sup>. Distances of 150–800 kpc for 10 of these objects have been estimated<sup>25</sup>, supporting a Local Group location. Kinematic models for the CHVCs have been constructed<sup>26</sup>, and have led to the conclusion that most CHVCs are at an average distance of 200 kpc from either our Galaxy or M31. CHVCs have also been viewed<sup>27</sup> as dark-matter-dominated ‘mini-haloes’ in orbit around the Galaxy (or M31) at characteristic distances of 150 kpc; in that model, a hot Galactic corona is required to pressure-confine the H I gas inside the mini-haloes.

On the basis of the above findings, and on the markedly different ionization state of the HV-O VI absorber presented here, and that of the isolated CHVCs, we suggest that CHVCs are the cold, condensed component of a multiphase local Inter-Galactic Medium in pressure equilibrium. The more diffuse and tenuous warm medium, possibly pervading the whole Local Group (and so spreading over distances of few megaparsecs), is responsible for the observed HV-O VI absorption, and the X-ray O VII, O VIII and Ne IX absorption systems, and provides the pressure needed to confine the cold phase found in the form of CHVCs. The hot plasma contains up to 100 times the total mass of the cold component<sup>24</sup>. The upper boundary of our baryonic mass estimate for the warm-hot phase suggests that this can account for up to 100% of the ‘missing’ baryons in our Local Group. □



**Figure 2** Velocity segregation of the HV-O VI absorbers. Aitoff plots of **a**, the LSR, and **b**, the Local Group Standard of Rest (LGSR) velocity distributions for the HV-O VI absorption systems. In **a** and **b**, filled circles correspond to negative velocities, and open circles correspond to positive velocities. The size of the circles is proportional to the amplitude of the velocity. In **a**, the open squares indicate the position of the barycentre of the Local Group (LG), M31 and the Virgo cluster, and the star marks the barycentre of the distribution. Red circles mark HV-O VI absorbers tentatively associated with HV-H I clouds with known distance, which have therefore been excluded from our statistical analysis.

**Table 1** The amplitude of the average HV-O VI velocity vector minimizes in the LGSR

| Reference frame | $\langle v \rangle (\text{km s}^{-1})$ | $\langle l \rangle (\text{deg.})$ | $\langle b \rangle (\text{deg.})$ |
|-----------------|----------------------------------------|-----------------------------------|-----------------------------------|
| LSR             | $121 \pm 17$                           | $277 \pm 12$                      | $33 \pm 27$                       |
| GSR             | $64 \pm 12$                            | $284 \pm 19$                      | $38 \pm 45$                       |
| LGSR            | $32 \pm 12$                            | $292 \pm 31$                      | $29 \pm 66$                       |

Symbols:  $v$ , velocity;  $l$ , Galactic longitude;  $b$ , Galactic latitude. Table shows average  $\langle v, l, b \rangle$  vectors, and associated uncertainties, for the high-velocity (HV)-O VI absorbers, in the Local Standard of Rest (LSR), the Galactic Standard of Rest (GSR) and the Local Group Standard of Rest (LGSR). Values are computed by: first, projecting  $(v, l, b)$  onto three cartesian axes  $(x, y, z)$ ; second, averaging the components and transforming  $\langle (x, y, z) \rangle$  back into  $\langle (v, l, b) \rangle$ ; and third, computing the standard deviation of the cartesian components, and propagating the errors in quadrature to obtain uncertainties on  $\langle (v, l, b) \rangle$ .

## Methods

### Data acquisition

We used only FUSE data from the LiF mirror and the A1 detector segment, which have optimal efficiency in the waveband around the O VI line. The data were calibrated and ‘cleaned’ following standard procedures<sup>28</sup>. The resolution of the final data is  $\sim 20 \text{ km s}^{-1}$ , at  $1,032 \text{ \AA}$ .

We selected for our sample only spectra with signal-to-noise ratio (S/N) per resolution element of  $>5$  at  $1,032 \text{ \AA}$  because we were interested in the O VI absorption lines. This selection produced 54 different lines of sight, all with  $|b| > 20^\circ$  (because of the high extinction suffered by lines of sight crossing a significant portion of the Galactic disk). In the lowest-quality spectra ( $S/N \approx 5$ ), absorption lines with equivalent width larger than  $\sim 200 \text{ m\AA}$  are detected at  $>3\sigma$ , whereas the highest S/N spectra in our sample guarantee the detection of absorption lines down to an equivalent width of  $\sim 20 \text{ m\AA}$ .

### Data analysis

For each  $z \approx 0$  O VI absorption line spectrum, we used the CIAO fitting engine Sherpa<sup>29</sup> to fit a power-law continuum plus negative gaussians to represent the absorption lines, and derived the best-fit wavelengths and widths over  $1,028.5 \text{ \AA}$  to  $1,034 \text{ \AA}$ .

We also looked for possible contamination by H<sub>2</sub> absorption lines from the Galactic interstellar medium, exploiting H<sub>2</sub> templates computed using a non-LTE (local thermodynamic equilibrium) photodissociation front code<sup>30,31</sup>. The excitation temperatures of the H<sub>2</sub> rotational–vibrational level distribution result from formation pumping, ultraviolet absorption of diffuse ultraviolet starlight in the electronic transitions to the states B, C, B' and D followed by fluorescence, quadrupole cascading, collisional excitation and de-excitation. Templates have been computed for broad ranges of kinetic temperatures, densities, Doppler parameters and factors of enhancing of the interstellar radiation field. Using these templates, we verified that, in the wavelength range of interest, spectra along lines of sight with equivalent hydrogen column densities lower than  $\sim 2 \times 10^{20} \text{ cm}^{-2}$  show only weak H<sub>2</sub> absorption. Spectra along lines of sight with  $N_{\text{H}} > 2 \times 10^{20} \text{ cm}^{-2}$ , instead, show usually strong (6–0) P(3)  $\lambda 1,031.19 \text{ \AA}$  and R(4)  $1,032.35 \text{ \AA}$  absorption. The P(3)  $\lambda 1,031.19 \text{ \AA}$  and R(4)  $1,032.35 \text{ \AA}$  lines occur at velocities of  $-214$  and  $+123 \text{ km s}^{-1}$ , and so may potentially contaminate HV-O VI lines. We found that about 30% of the negative HV-O VI lines and less than 20% of the positive HV-O VI lines had profiles contaminated by the P(3)  $\lambda 1,031.19 \text{ \AA}$  and R(4)  $1,032.35 \text{ \AA}$  lines, respectively. For all spectra we used other strong H<sub>2</sub> lines in otherwise relatively featureless portions of the FUSE spectra, to determine the bulk velocity of the H<sub>2</sub> absorber and the strengths of the P(3)  $\lambda 1,031.19 \text{ \AA}$  and R(4)  $1,032.35 \text{ \AA}$  lines from the relative oscillator strength ratios, allowing us to disentangle the HV-O VI line from the contaminating H<sub>2</sub>. In particular, we used the (6–0) P(2)  $\lambda 1,028.10 \text{ \AA}$ , (6–0) R(3)  $\lambda 1,028.98 \text{ \AA}$  and (8–0) P(4)  $\lambda 1,012.26 \text{ \AA}$  lines, on the short wavelength side of the O VI  $\lambda 1,031.926 \text{ \AA}$  line, and the (5–0) P(2)  $\lambda 1,040.36 \text{ \AA}$ , (5–0) R(3)  $\lambda 1,041.16 \text{ \AA}$  and (5–0) R(4)  $\lambda 1,044.54 \text{ \AA}$  lines, on the long wavelength side. Only in two cases could the presence of the HV-O VI line not be safely established, and so we considered those two lines of sight as free from HV-O VI absorption, at our detection threshold. In three more cases the presence of the HV-O VI line was clearly established, but its position and width only poorly constrained.

### Cross-correlation

To identify HV-O VI lines with possible HV-H I counterparts, we cross-correlated the objects of our sample with catalogues of 21-cm H I lines<sup>17,18</sup>. We used very conservative searching criteria, and define a ‘coincidence’ when Galactic coordinates and LSR velocities of the HV-O VI absorbers fall into the very broad ranges (typically tens of degrees and about  $100 \text{ km s}^{-1}$ ) listed in the HV-H I catalogues. The very broad range in coordinates depends, not on the beam-size of the telescope with which 21-cm measurements were performed, but on the entire size of the group or complex of HV-H I clouds to which a given cloud is supposed to belong. This search gave five coincidences. For completeness, we also cross-correlated our coincidences with those found<sup>4</sup> in the compilation of HV-O VI absorbers<sup>16</sup>. This search left only one object. This object has been tentatively associated<sup>16</sup> with the Magellanic stream HV-H I complex, although not with any particular cloud. To be conservative we then add this object to the list of possible OVI-H I identification, and took these six objects off our sample for the purpose of the presented statistical analysis.

Received 16 May; accepted 13 December 2002; doi:10.1038/nature01369.

1. Cen, R. & Ostriker, J. P. Where are the baryons? *Astrophys. J.* **514**, 1–6 (1999).
2. Hellsten, U., Gnedin, N. Y. & Miralda-Escudé, J. The X-ray forest: A new prediction of hierarchical structure formation models. *Astrophys. J.* **509**, 56–61 (1998).
3. Davé, R. *et al.* Baryons in the warm-hot intergalactic medium. *Astrophys. J.* **552**, 473–483 (2001).
4. Sembach, K. R. *et al.* Highly-ionized high-velocity gas in the vicinity of the galaxy. *Astrophys. J. Suppl.* (2002) (submitted); preprint astro-ph/0207562 at (<http://xxx.lanl.gov>) (2002).
5. Kahn, F. D. & Woltjer, L. Intergalactic matter and the galaxy. *Astrophys. J.* **130**, 705–717 (1959).
6. Nicastro, F. *et al.* Chandra discovery of a tree in the X-ray forest towards PKS 2155–304: The local filament? *Astrophys. J.* **573**, 157–167 (2002).
7. Canizares, C. R., Schattenburg, M. L. & Smith, H. I. The high energy transmission grating spectrometer for AXAF. *Proc. SPIE* **597**, 253–260 (1986).
8. Moos, H. W. *et al.* Overview of the Far Ultraviolet Spectroscopic Explorer mission. *Astrophys. J.* **538**, L1–L6 (2000).
9. Savage, B. D. *et al.* Far Ultraviolet Spectroscopic Explorer observations of O VI absorption in the galactic halo. *Astrophys. J.* **538**, L27–L30 (2000).
10. Wakker, B. P. *et al.* The FUSE survey of galactic OVI. *Am. Astron. Soc.* **199**, 6508 (2001).
11. Savage, *et al.* Distribution and kinematics of OVI in the Milky Way halo. *Am. Astron. Soc.* **199**, 6504 (2001).
12. Howk, J. C., Savage, B. D., Sembach, K. R. & Hoopes, C. G. Far-Ultraviolet Spectroscopic Explorer observations of degree-scale variations in galactic halo OVI. *Am. Astron. Soc.* **572**, 264–275 (2002).

13. Hoopes, C. G., Sembach, K. R., Howk, J. C., Savage, B. D. & Fullerton, A. W. A Far Ultraviolet Spectroscopic Explorer survey of interstellar O VI absorption in the Small Magellanic Cloud. *Astrophys. J.* **569**, 233–244 (2002).
14. Sembach, K. R. *et al.* Far Ultraviolet Spectroscopic Explorer observations of O VI in high-velocity clouds. *Astrophys. J.* **538**, L31–L34 (2000).
15. Sembach, K. R. High velocity clouds: Cosmological and galactic weather. *Am. Astron. Soc.* **199**, 11301 (2001).
16. Heckman, T. M., Norman, C. A., Strickland, D. K. & Sembach, K. R. On the physical origin of OVI absorption-line systems. *Astrophys. J.* (in the press); preprint astro-ph/020556 at (<http://xxx.lanl.gov>) (2002).
17. Stark, A. A. *et al.* The Bell Laboratories H I survey. *Astrophys. J. Suppl. Ser.* **79**, 77–104 (1992).
18. Wakker, B. P. Distances and metallicities of high- and intermediate-velocity clouds. *Astrophys. J. Suppl. Ser.* **136**, 463–535 (2001).
19. Oort, J. H. Asymmetry in the distribution of stellar velocities. *Observatory* **49**, 302–304 (1926).
20. Kravtsov, A. V., Klypin, A. & Hoffman, Y. Constrained simulations of the real universe: II. Observational signatures of intergalactic gas in the local supercluster region. *Astrophys. J.* (in the press); preprint astro-ph/0109077 at (<http://xxx.lanl.gov>) (2001).
21. Klypin, A., Hoffman, Y., Kravtsov, A. & Gottloeber, S. Constrained simulations of the real universe: The local supercluster. *Astrophys. J.* (2001) (submitted); preprint astro-ph/0107104 at (<http://xxx.lanl.gov>) (2001).
22. Grevesse, N. & Anders, E. *Cosmic Abundances of Matter* (ed. Waddington, C. J.) 1–8 (AIP Conference Proceedings Vol. 183, American Institute of Physics, New York, 1989).
23. Blitz, L., Spergel, D. N., Teuben, P. J., Hartmann, D. & Burton, W. B. High-velocity clouds: Building blocks of the local group. *Astrophys. J.* **514**, 818–843 (1999).
24. Braun, R. & Burton, W. B. The kinematic and spatial deployment of compact, isolated high-velocity clouds. *Astrophys. J.* **341**, 437–450 (1999).
25. Braun, R. & Burton, W. B. Status of HI searches for CHVCs beyond the local group. *Astron. Astrophys.* **375**, 219–226 (2001).
26. de Heij, V., Braun, R. & Burton, W. B. An all-sky study of compact isolated high velocity clouds. *Astron. Astrophys.* (in the press); preprint astro-ph/0206306 at (<http://xxx.lanl.gov>) (2002).
27. Sternberg, A., McKee, C. F. & Wolfire, M. G. Atomic hydrogen gas in dark-matter minihalos and the compact high velocity clouds. *Astrophys. J. Suppl.* (2002) (in the press); preprint astro-ph/0207040 at (<http://xxx.lanl.gov>) (2002).
28. Sankrit, R., Kruk, J. W., Ake, T. B. & Anderson B.-G. *The Fuse Data Analysis Cookbook* v. 1.0 (<http://fuse.pha.jhu.edu/analysis.cookbook.html>) (2001).
29. Freeman, P., Doe, S. & Siemiginowska, A. Sherpa: A mission-independent data analysis application. *Proc. SPIE* **4477**, 76–87 (2001).
30. Draine, B. T. & Bertoldi, F. Structure of stationary photodissociation fronts. *Astrophys. J.* **468**, 269–289 (1996).
31. Cecchi-Pestellini, C., Casu, S. & Dalgarno, A. The population of the high H<sub>2</sub> rotational states towards the OMC-1 peak 1. *Astrophys. J.* (submitted).

**Acknowledgements** F.N. thanks P. Kaaret and G. C. Perola for discussions and comments. We thank W.B. Burton for comments that helped improve the Letter. This work was partly supported by NASA-Chandra grants and a NASA-Chandra X-ray Center contract.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to F.N. (e-mail: [fnicastro@cfa.harvard.edu](mailto:fnicastro@cfa.harvard.edu)).

## Experimental realization of freely propagating teleported qubits

Jian-Wei Pan, Sara Gasparoni, Markus Aspelmeyer, Thomas Jennewein & Anton Zeilinger

Institut für Experimentalphysik, Universität Wien, Boltzmanngasse 5, 1090 Wien, Austria

Quantum teleportation<sup>1</sup> is central to quantum communication, and plays an important role in a number of quantum computation protocols<sup>2,3</sup>. Most information-processing applications of quantum teleportation include the subsequent manipulation of the qubit (the teleported photon), so it is highly desirable to have a teleportation procedure resulting in high-quality, freely flying qubits. In our previous teleportation experiment<sup>4</sup>, the teleported qubit had to be detected (and thus destroyed) to verify the success of the procedure. Here we report a teleportation experiment that results in freely propagating individual qubits. The basic idea is to suppress unwanted coincidence detection events by providing



the photon to be teleported much less frequently than the auxiliary entangled pair. Therefore, a case of successful teleportation can be identified with high probability without the need actually to detect the teleported photon. The experimental fidelity of our procedure surpasses the theoretical limit required for the implementation of quantum repeaters<sup>5,6</sup>.

To see how to achieve free propagation of teleported qubits, we first briefly review the working principle of the earlier Innsbruck teleportation experiment as illustrated in Fig. 1a. Suppose photon 1, which Alice wants to teleport to Bob, is in a general polarization state  $|\Phi\rangle_1 = \alpha|H\rangle_1 + \beta|V\rangle_1$  (unknown to Alice), and the pair of photons 2 and 3 shared by Alice and Bob is in the polarization entangled state  $|\Phi^+\rangle_{23}$ . We use the usual Bell states:

$$|\Phi^\pm\rangle_{ij} = \frac{1}{\sqrt{2}}(|H\rangle_i|H\rangle_j \pm |V\rangle_i|V\rangle_j)$$

$$|\Psi^\pm\rangle_{ij} = \frac{1}{\sqrt{2}}(|H\rangle_i|V\rangle_j \pm |V\rangle_i|H\rangle_j) \quad (1)$$

where  $H$  and  $V$  denote horizontal and vertical linear polarizations, and  $i$  and  $j$  index the spatial modes of the photons. We choose the state  $|\Phi^+\rangle_{23}$  just for the convenience of discussion. The overall state of photons 1, 2 and 3 can then be rewritten as:

$$|\Psi\rangle_{123} = |\Phi\rangle_1 |\Phi^+\rangle_{23}$$

$$= \frac{1}{2} [|\Phi^+\rangle_{12}(\alpha|H\rangle_3 + \beta|V\rangle_3) + |\Phi^-\rangle_{12}(\alpha|H\rangle_3 - \beta|V\rangle_3)$$

$$+ |\Psi^+\rangle_{12}(\alpha|V\rangle_3 + \beta|H\rangle_3) + |\Psi^-\rangle_{12}(\alpha|V\rangle_3 - \beta|H\rangle_3)] \quad (2)$$

It can thus be seen that a joint Bell measurement on photons 1 and 2 at Alice's side projects the state of photon 3 at Bob's side into one of the four corresponding states as shown in equation (1). Depending on Alice's measurement results, Bob can then perform a unitary transformation, independent of  $|\Phi\rangle_1$ , on photon 3 to convert its state into the initial state of photon 1. This is how quantum teleportation works. Obviously, to demonstrate the working principle of teleportation it is sufficient to identify one of the four Bell-states. This results in a reduced efficiency—the fraction of success—of 25%. In the Innsbruck experiment, this is accomplished by detecting a twofold coincidence behind a 50:50 beamsplitter.

So far, we considered only the ideal case where there is one and only one photon pair in each of the pairs of modes 1-4 and 2-3. Then it is clear that a threefold coincidence of detectors D1-D2-T is enough to guarantee a successful teleportation. However, because of the probabilistic feature of pair creation in spontaneous parametric down-conversion (SPDC), this conclusion cannot be applied to the Innsbruck experiment.

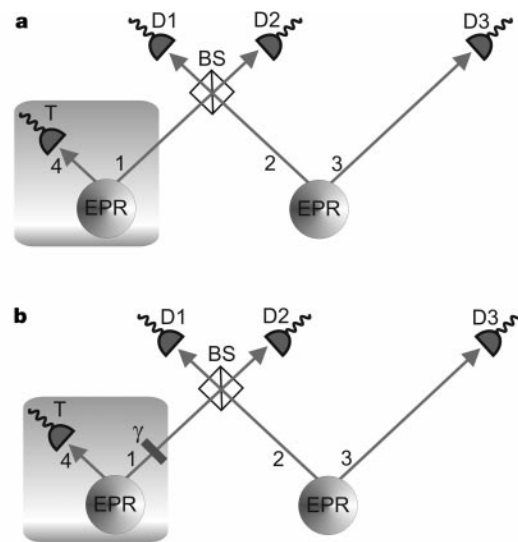
To see the problem more clearly, let us denote by  $p$  the probability of having a single pair creation during a SPDC process. For independent probabilistic pair creation the probability of having one photon pair in each of the pairs of modes 1-4 and 2-3, which corresponds to a successful teleportation event, is then given by  $p^2$ . Yet sometimes two pairs of photons are both emitted into either modes 1-4 or 2-3. Specifically, with the probability of  $p^2$  the two photon pairs could both appear in modes 1-4 while no photons are present in modes 2-3. In this case, although a threefold coincidence of D1, D2 and T is detected, no teleportation occurs because of the absence of the entangled pair in modes 2-3 and mode 3 is simply empty. To ensure a successful teleportation, it is therefore necessary in the Innsbruck experiment to confirm the presence of photon 3 by actually detecting it. For this reason, the Innsbruck experiment has been called a "postselected" one<sup>7</sup>; using the word "conditional" would be more appropriate as photon 3 is not selected depending on its state<sup>8</sup>.

Most applications of quantum teleportation include the subsequent manipulation of the teleported photon. Thus, a freely

propagating output state, which is teleported with high fidelity, is strongly desired. This calls for a non-conditional qubit teleportation scheme. Possible solutions<sup>7,9</sup> could include the discrimination of one- and two-photon events at detector T, a quantum non-demolition measurement of the photon number in mode 3, or the introduction of an unbalanced coupling between modes 1-4 relative to modes 2-3. But owing to the lack of appropriate technology, none of the proposed schemes has so far been realized.

In other experimental efforts<sup>10,11</sup> to achieve full teleportation, other problems still remain for various reasons. Concerning the case of continuous variable teleportation<sup>10</sup>, criticism has been raised due to the low fidelity<sup>12</sup> and the need of a phase reference<sup>13</sup>. In the other case<sup>11</sup>, the input state fails to meet the criterion of being a quantum state, since about  $10^5$  photons are prepared in the input beam.

In the present experiment, we achieve a freely propagating teleported quantum state by making a detection at Bob's side obsolete (Fig. 1b). The main point is to reduce the number of unwanted D1-D2 coincidence counts. This is accomplished by attenuating the beam 1 by a factor of  $\gamma$  while leaving the other modes unchanged. Then a threefold coincidence D1-D2-T will occur with probability  $\gamma p^2$  owing to successful teleportation and with significantly lower probability  $(\gamma p)^2$  they will be a spurious coincidence. Thus, for sufficiently low  $\gamma$  it will no longer be necessary to actually detect the teleported photon 3 and a freely propagating teleported beam of qubits emerges. In essence, we provide the entangled ancillary pair more frequently than the photon to be teleported. Thus it can only very rarely happen that



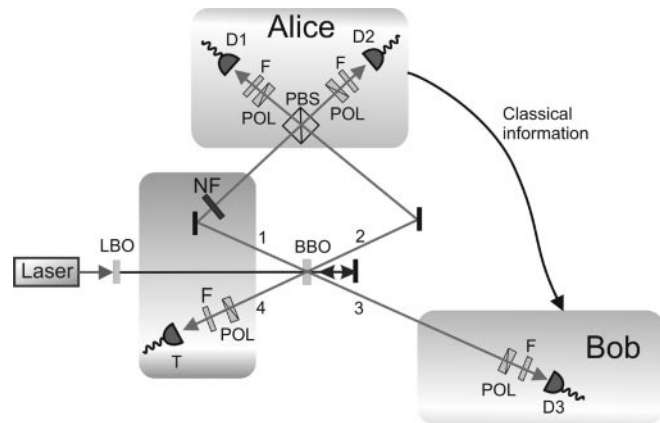
**Figure 1** Diagrams showing the principles of the Innsbruck experiment and of achieving free propagation of teleported qubits. **a**, The Innsbruck experiment. Two independent polarization entangled photon pairs produced by spontaneous parametric down-conversion (SPDC) are used to achieve quantum teleportation. While the first pair (photons 2 and 3) provides the entangled pair needed in the teleportation protocol, the second pair (photons 1 and 4) is used to prepare the initial state to be teleported. The 'twin' photon 4 acts only as a trigger. In the case where one pair of photons is emitted in each of the pairs of modes 1-4 and 2-3, a threefold coincidence of T-D1-D2 is enough to indicate a successful teleportation. However, in the case where both photon pairs are emitted into modes 1-4, threefold coincidences of T-D1-D2 also result. In that case, no teleportation occurs because mode 3 is simply empty. Therefore detection of the teleported photon 3 is needed to rule out this bad case. **b**, Free propagation of teleported qubits. An additional attenuator is inserted into mode 1 to decrease the probability of unwanted, spurious threefold coincidence detection events of D1-D2-T resulting from double-pair emission into modes 1-4. This method makes a detection at Bob's side obsolete and leads to free propagation of teleported qubits (see text). EPR, Einstein-Podolsky-Rosen entangled photon sources. BS, beamsplitter. D1, D2 and T, detectors.

the teleportation machinery is not ready when a photon arrives to be teleported.

A schematic drawing of the experimental set-up is shown in Fig. 2. As in our recent experimental demonstration of four-photon entanglement<sup>14</sup>, we exploit a polarizing beamsplitter (PBS) instead of a normal beamsplitter to perform the necessary Bell-state measurement. In the experiment, we prepared the entangled pair in state  $|\Phi^+\rangle_{23}$  and chose to analyse the projection onto state  $|\Phi^+\rangle_{12}$ . This projection can be accomplished by performing a joint polarization measurement in the  $45^\circ$  basis behind the PBS. As stated in ref. 15, registering  $+45^\circ/+45^\circ$  or  $-45^\circ/-45^\circ$  coincidences of detectors D1 and D2 acts as a projection onto  $|\Phi^+\rangle_{12}$ . This projection leaves photon 3 in the initial state (see equation (2)).

Here we emphasize that one can neglect the cases where a single pair production in modes 1-4 and a double pair production in modes 2-3 give rise to third-order spurious threefold coincidences with probability  $p^3$ . This is because in our experiment  $\gamma \gg p$ . Therefore, such third-order spurious coincidences are at most one-tenth of those contributed from the double pair emission in modes 1-4, and are even much smaller than the good coincidence events. In that sense, our scheme to attenuate the beam 1 is equivalent to the one to attenuate the reflected pump beam as suggested in ref. 9.

A high-intensity source of entangled photon pairs is desirable to collect enough experimental data within a reasonably short time, because the method of beam attenuation significantly reduces the overall fourfold coincidences (typically by one order of magnitude). Various efforts have been made to achieve this purpose. By using a compact set-up and by focusing the ultraviolet (UV) pump (with a beam diameter  $<0.5$  mm) onto the beta-barium borate (BBO) crystal, we achieved both better stability and collection efficiency. We thus managed to observe about  $2 \times 10^4$  entangled photon pairs per second, while achieving a high twofold visibility of 94% in the  $45^\circ$  basis. Correspondingly, the observed overall fourfold coincidence rate was as high as two per second on average, which is almost two orders of magnitude higher than recent work in quantum teleportation<sup>14,16</sup>. Note that high multi-fold coincidences have



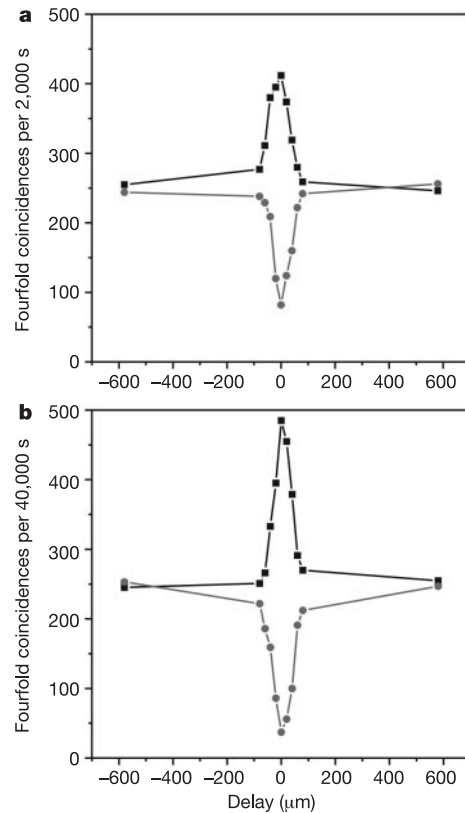
**Figure 2** Set-up for experimental demonstration of free propagation of teleported qubits. A mode-locked laser produces pulses of 200 fs with a repetition rate of 76 MHz, centred on the wavelength of 788 nm. This radiation is used to pump an up-conversion process in an LBO and the outgoing radiation is further sent to a BBO crystal. Two independent EPR pairs are produced through SPDC, the first during the direct passage of the pump light (2 and 3), and the second during its back propagation (1 and 4). A polarizer (POL) is put in front of the trigger detector T to prepare photon 1 nonlocally in the state to be teleported. Various neutral filters (NF) are inserted into beam 1 to achieve the required attenuation. Photons 1 and 2 are then steered to a polarizing beamsplitter (PBS) to perform the necessary Bell-state measurement. To ensure good temporal and spatial overlap of the two photons, the two outputs of the PBS are spectrally filtered (3-nm bandwidth) and monitored by fibre-coupled single photon detectors. LBO, lithium triborate crystal. F, narrow bandwidth filters.

also been observed in very different experiments<sup>17,18</sup>.

To experimentally demonstrate non-conditional teleportation, we inserted a series of neutral filters in mode 1 and showed that the probability of having a successful teleportation dependent on a threefold coincidence of D1, D2 and T increases with decreasing filter transmission  $\gamma$ .

First, we performed teleportation of a  $45^\circ$  polarized photon. The state to be teleported was prepared nonlocally by using the entanglement of photons 1 and 4. In the experiment, both the threefold coincidences  $C_3$  of D1, D2 and T and the fourfold coincidences  $C_4$  of all outputs were measured. As an example, the measurement results for the case of  $\gamma = 1$  (that is, no attenuation) is shown in Fig. 3a. It is obvious that the observed probability to have a successful teleportation once a threefold coincidence occurs is given by  $C_4/C_3$ . Owing to the finite detection efficiency  $\eta_3$  of D3, only a fraction of the total successful teleportation events will be registered by detector D3. This implies that the genuine probability of success is given by the observed probability of success divided by the detection efficiency, that is  $C_4/(C_3\eta_3)$ .

Measuring  $C_3$ ,  $C_4$  and  $\eta_3$  (typically 0.15), we can thus obtain the observed and genuine probabilities to have a successful teleportation, based on a threefold coincidence, for different neutral filters. Table 1 summarizes the experimental results in our  $45^\circ$  teleportation experiment. The experimental results well confirm,



**Figure 3** Two typical experimental results for  $45^\circ$  teleportation. **a**, Fourfold coincidences between the detectors T, D1, D2 and D3 as a function of time delay between the arrival of photons 1 and 2 at the PBS. Shown are the cases where the polarizer in front of D3 is set to  $45^\circ$  (square) and  $-45^\circ$  (circle). At large delay, no two-photon interference between photons 1 and 2 occurs, and therefore the polarization correlation between photons 1 and 3 is completely random. At zero delay a perfect overlap between photons 1 and 2 leads to a successful projection onto  $|\Phi^+\rangle_{12}$  and results in teleportation of the input state  $|\Phi\rangle_1$ . The data shown were measured without using neutral filters and before birefringence compensation. **b**, The experimental visibility is drastically enhanced from 66% to 86% after birefringence compensation. Note that in this measurement, one neutral filter with  $\gamma = 0.05$  was used just to obtain the unbalancing.

Table 1 Probabilities of success for different filters

| Attenuation | Observed          | Genuine         |
|-------------|-------------------|-----------------|
| 1           | $0.060 \pm 0.001$ | $0.40 \pm 0.01$ |
| 0.50        | $0.079 \pm 0.001$ | $0.53 \pm 0.01$ |
| 0.10        | $0.119 \pm 0.002$ | $0.79 \pm 0.01$ |
| 0.05        | $0.138 \pm 0.002$ | $0.92 \pm 0.01$ |

in agreement with our above analysis, that the probability of success increases with decreasing  $\gamma$ . Note that for a given  $\gamma$  all data points in scan (Fig. 3) are used to estimate the probability of success, which results in a small statistical error.

Using the product of the measured fidelities and probabilities of a successful teleportation, we can obtain the non-conditional fidelity. The corresponding fidelities are shown in Fig. 4 for the conditional detection scheme and for the new unbalanced interferometer scheme at different grades of attenuation. If no attenuation is used the set-up is in principle equivalent to the original Innsbruck scheme and the freely propagating output state fully incorporates the vacuum contribution, resulting in a non-conditional fidelity below the classical limit. With increasing attenuation ('unbalancing') the vacuum contribution can be suppressed, resulting in a higher non-conditional fidelity. For the case where an attenuation factor of  $\gamma = 0.05$  is chosen, the experimental non-conditional fidelity is  $0.79 \pm 0.02$ , which significantly overcomes the classical limit of two-thirds.

For a complete demonstration of our teleportation method, in the experiment we also teleported photons with horizontal linear polarization (that is,  $H$ ) and right-handed circular polarization (that is,  $R$ ), thus demonstrating the scheme for three orthogonal directions on the Poincaré sphere. The non-conditional fidelities measured for different preparation bases are summarized in Table 2. We conclude that in all three cases the obtained non-conditional fidelities are well above the classical limit. We noticed that, whereas both the  $45^\circ$  and  $R$  teleportation experiments have roughly the same fidelities, a much better fidelity is obtained in the  $H$  teleportation experiment. This is because we chose the  $H$ - $V$  direction as the intrinsic measurement basis of the polarizing beamsplitter cube and so the direction of polarization is aligned with respect to the optical axes. All input pairs of  $H$  or  $V$  polarization only collect an overall phase and are thus, in the case of projection onto a  $|\Phi^+\rangle$ -state, always directed into different beamsplitter outputs. For the case of arbitrary polarized

Table 2 Fidelities of non-conditional teleportation

| Polarization | $\gamma = 1$    | $\gamma = 0.05$ |
|--------------|-----------------|-----------------|
| $45^\circ$   | $0.35 \pm 0.02$ | $0.79 \pm 0.02$ |
| $H$          | $0.39 \pm 0.02$ | $0.84 \pm 0.02$ |
| $R$          | $0.34 \pm 0.02$ | $0.76 \pm 0.02$ |

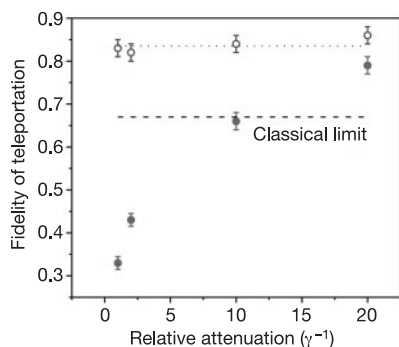
input states, an additional phase difference between horizontal and vertical components is introduced in the  $|\Phi\rangle$ -states, which consequently leads to a reduction of teleportation fidelity.

However, we can introduce an additional birefringent element in order to compensate for the above effect. After this compensation, as a test, we repeated the  $45^\circ$  teleportation experiment for the cases of  $\gamma = 1$  and  $\gamma = 0.05$ . The experimental result for  $\gamma = 0.05$  is shown in Fig. 3b. The compensation of birefringence effects in the experimental set-up increases the non-conditional fidelity in the  $45^\circ$  teleportation from  $0.79 \pm 0.02$  to  $0.85 \pm 0.02$ , which is compatible with that obtained in our  $H$  teleportation. This improvement implies that the PBS accuracy of interfering two independent photons has been raised from 89% to 97%, which is the fidelity of the Bell-state measurement of  $|\Phi^+\rangle$ . This is obtained from the experimental visibility for  $45^\circ$  teleportation after taking into account the imperfection of the state preparation. We emphasize that such a high accuracy achieved in the present experiment now fulfills the strict precision requirements of 95% for local operations of independent photons necessary for quantum repeaters in long-distance quantum communication<sup>5,6</sup>.

We were thus able to demonstrate the preparation of a freely propagating teleported quantum state with high (non-conditional) fidelity of  $0.85 \pm 0.02$ , that is, well above the classical limit. Our present results also confirm that the Innsbruck experiment was a bona fide realization of quantum teleportation. Furthermore, together with the expected future realization of entanglement purification<sup>19,20</sup>, the high experimental visibility obtained will allow the next step towards long-distance quantum communication.  $\square$

Received 26 September; accepted 30 December 2002; doi:10.1038/nature01412.

- Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **83**, 3081–3084 (1993).
- Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
- Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
- Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
- Briegleb, H.-J., Duer, W., Cirac, J. I. & Zoller, P. Quantum repeaters: The role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
- Aschauer, H. & Briegel, H. J. Private entanglement over arbitrary distances, even using noisy apparatus. *Phys. Rev. Lett.* **88**, 047902 (2002).
- Braunstein, S. L. & Kimble, H. J. *A posteriori* teleportation. *Nature* **394**, 840–841 (1998).
- Bouwmeester, D. *et al.* Reply to "A posteriori teleportation". *Nature* **394**, 841 (1998).
- Kok, P. & Braunstein, S. L. Postselected versus nonpostselected quantum teleportation using parametric down-conversion. *Phys. Rev. A* **61**, 042304 (2000).
- Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
- Kim, Y. H., Kulik, S. P. & Shih, Y. H. Quantum teleportation of a polarization state with a complete Bell state measurement. *Phys. Rev. Lett.* **86**, 1370–1373 (2001).
- Grosshans, F. & Grangier, P. Quantum cloning and teleportation criteria for continuous quantum variables. *Phys. Rev. A* **64**, R010301 (2001).
- Rudolph, T. & Sanders, B. C. Requirement of optical coherence for continuous variable quantum teleportation. *Phys. Rev. Lett.* **88**, 077903 (2001).
- Pan, J.-W. *et al.* Experimental demonstration of four-photon entanglement and high-fidelity teleportation. *Phys. Rev. Lett.* **86**, 4435–4438 (2001).
- Pan, J.-W. & Zeilinger, A. Greenberger-Horne-Zeilinger-state analyzer. *Phys. Rev. A* **57**, 2208–2211 (1998).
- Jennewein, T., Weihs, G., Pan, J.-W. & Zeilinger, A. Experimental nonlocality proof of quantum teleportation and entanglement swapping. *Phys. Rev. Lett.* **88**, 017093 (2002).
- Kurtsiefer, C., Oberparleiter, M. & Weinfurter, H. High-efficiency entangled photon pair collection in type-II parametric fluorescence. *Phys. Rev. A* **64**, 023802 (2001).
- Howell, J. C., Lamas-Linares, A. & Bouwmeester, D. Experimental violation of a spin-1 Bell inequality using maximally entangled four-photon states. *Phys. Rev. Lett.* **88**, 030401 (2002).
- Bennett, C. H. *et al.* Purification of noisy entanglement, and faithful teleportation via noisy channel. *Phys. Rev. Lett.* **76**, 1895–1898 (1996).
- Pan, J.-W., Simon, C., Brukner, C. & Zeilinger, A. Entanglement purification for quantum communication. *Nature* **410**, 1067–1070 (2001).



**Figure 4** Conditional fidelities and non-conditional fidelities obtained in  $45^\circ$  teleportation for different neutral filters. With increasing attenuation of the beam 1, we observe an increase of non-conditional fidelities (black symbols) while the conditional ones (open symbols) remain constant. In the experiment, the integration times for various  $\gamma$  were, respectively, about  $2 \times 10^3$ ,  $4 \times 10^3$ ,  $2 \times 10^4$  and  $4 \times 10^4$  s per measurement point on the average. For each  $\gamma$  we collected about 400 fourfold counts for peak and 80 for dip. For  $\gamma = 0.10$ , the classical limit is already reached, and for  $\gamma = 0.05$  it is clearly overcome, resulting in freely propagating teleported photon states of high fidelity.



**Acknowledgements** This work was supported by the Austrian Science Foundation (FWF), the TMR and the QuComm programmes of the European Commission and the Alexander von Humboldt Foundation.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to J.-W.P. (e-mail: pan@ap.univie.ac.at) or A.Z. (e-mail: Zeilinger-office@exp.univie.ac.at).

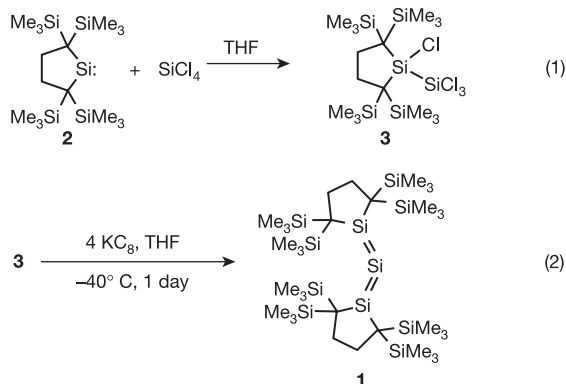
## A stable silicon-based allene analogue with a formally *sp*-hybridized silicon atom

S. Ishida, T. Iwamoto, C. Kabuto & M. Kira

Department of Chemistry, Graduate School of Science, Tohoku University, Aoba-ku, Sendai 980-8578, Japan

Carbon chemistry exhibits a rich variety in bonding patterns, with homo- or heteronuclear multiple bonds involving *sp*-hybridized carbon atoms as found in molecules such as acetylenes, nitriles, allenes and carbon dioxide. Carbon's heavier homologues in group 14 of the periodic table—including silicon, germanium and tin—were long thought incapable of forming multiple bonds because of the less effective  $p\pi-p\pi$  orbital overlap involved in the multiple bonds. However, bulky substituents can protect unsaturated bonds and stabilize compounds with formally *sp*-hybridized heavy group-14 atoms<sup>1,2</sup>: stable germanium<sup>2</sup>, tin<sup>3</sup> and lead<sup>4</sup> analogues of acetylene derivatives and a marginally stable tristannaallene<sup>5</sup> have now been reported. However, no stable silicon compounds with formal *sp*-silicon atoms have been isolated. Evidence for the existence of a persistent disilaacetylene<sup>6</sup> and trapping<sup>7</sup> of transient 2-silaallenes and other  $X=Si=X'$  type compounds ( $X, X' = O, CR_2, NR$ , and so on) are also known, but stable silicon compounds with formally *sp*-hybridized silicon atoms have not yet been isolated. Here we report the synthesis of a thermally stable, crystalline trisilaallene derivative containing a formally *sp*-hybridized silicon atom. We find that, in contrast to linear carbon allenes, the trisilaallene is significantly bent. The central silicon in the molecule is dynamically disordered, which we ascribe to ready rotation of the central silicon atom around the molecular axis.

Trisilaallene **1** was prepared as a green solid using two-step reactions from silylene **2** (ref. 8) in overall 42% yield (equations (1) and (2)). Trisilaallene **1** is sensitive to air but is thermally rather

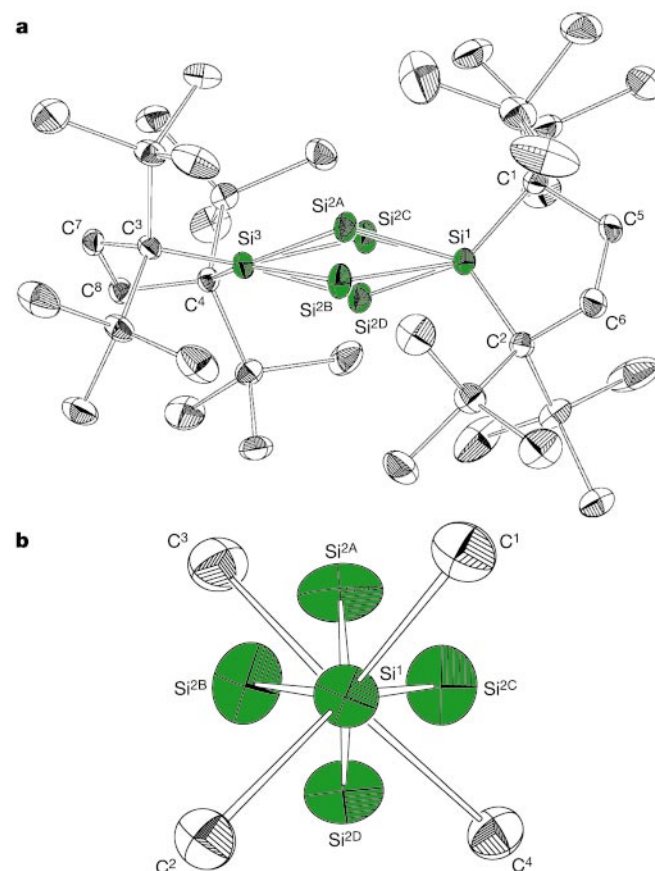


stable with a melting point of 198–200 °C. The structure of **1** was determined by <sup>1</sup>H, <sup>13</sup>C, and <sup>29</sup>Si NMR spectroscopy and X-ray

crystallography (see Supplementary Information for details).

X-ray single crystal analysis shows a quite unusual structural feature of trisilaallene **1** (Fig. 1). In contrast to carbon allenes, the trisilaallene skeleton is not linear but bent. The central silicon atom (Si<sup>2</sup>) in the crystal was found at four positions labelled Si<sup>2A</sup> to Si<sup>2D</sup> in Fig. 1a at higher temperatures than -100 °C, indicating that four structurally similar isomers exist, A to D. The populations for A to D are independent of crystals but are significantly temperature-dependent: the populations are 46, 23, 22 and 10% at 0 °C, but 76, 18, 7 and 0% at -150 °C, respectively (see Supplementary Information for details). The energy differences between isomers A to D are estimated to be within 1 kcal mol<sup>-1</sup>, suggesting a dynamic disorder<sup>9</sup> mediated by a rotation of the Si<sup>2</sup> atom around the Si<sup>1</sup>–Si<sup>3</sup> axis.

The details of the molecular structure of isomer A of trisilaallene **1** at -150 °C are as follows: (1) The Si<sup>1</sup>=Si<sup>2</sup>=Si<sup>3</sup> skeleton is not linear but is significantly bent with a bond angle of 136.49(6)° (standard deviation is in parentheses), indicating that the bonding at Si<sup>2</sup> atom cannot be described by using simple *sp*-hybridization any longer. (2) The two Si=Si bond lengths in **1** (2.177(1) and 2.188(1) Å) are in the range of those for typical stable disilenes (2.14–2.29 Å). (3) The two five-membered rings are almost perpen-

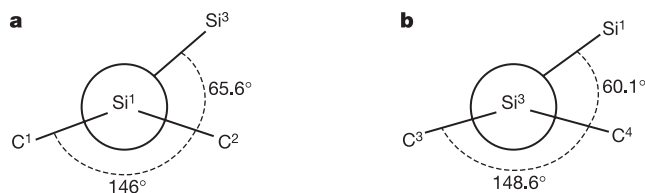


**Figure 1** Molecular structure of trisilaallene **1** at -50 °C. **a**, A top view (hydrogen atoms were omitted for clarity). **b**, A side view (C<sub>2</sub>Si<sup>1</sup>=Si<sup>2</sup>=Si<sup>3</sup>C<sub>2</sub> skeleton). As shown in **b**, Si<sup>2A</sup>–Si<sup>2D</sup> are found in four quadrants divided by two perpendicular planes of C<sup>1</sup>–Si<sup>1</sup>–C<sup>2</sup> and C<sup>3</sup>–Si<sup>3</sup>–C<sup>4</sup>, respectively. The environments of the four quadrants are slightly different from each other owing to the non-planarity of the five-membered rings, and hence, the occupancies of Si<sup>2A</sup>–Si<sup>2D</sup> are not equal to each other. X-ray analysis for two different single crystals of **1** showed that the populations are independent of the crystals but are significantly temperature-dependent. The populations of isomers A to D are: 53, 22, 19 and 7% at -50 °C and 76, 18, 7 and 0% at -150 °C for crystal 1; 46, 23, 22 and 10% at 0 °C; 52, 22, 19 and 7% at -50 °C; 61, 21, 15 and 4% at -100 °C; and 72, 19, 10 and 0% at -146 °C for crystal 2.

dicular to each other; the dihedral angle between the  $C^1-Si^1-C^2$  and  $C^3-Si^3-C^4$  planes is  $92.5^\circ$ . (4) The geometry around the two terminal silicon atoms is not planar but pyramidalized (the sums of the bond angles around  $Si^1$  and  $Si^3$  atoms are  $354.1^\circ$  and  $354.9^\circ$ , respectively). (5) As shown in Fig. 2, the dihedral angles  $Si^3-Si^2-Si^1-C^2$  and  $Si^1-Si^2-Si^3-C^4$  are  $65.6(1)$  and  $60.1(2)^\circ$ , respectively, indicating a *trans*-bent arrangement around  $Si^1-Si^2$  and  $Si^3-Si^2$  bonds. All these structural features indicate that trisilaallene **1** is characterized as the first compound with two *trans*-bent  $Si=Si$  double bonds that share a silicon atom. A similar structure was found for a trisilallene synthesized by ref. 5, whereas the skeleton was depicted by a different bonding scheme.

The *ab initio* molecular orbital calculations for a model trisilaallene  $(H_3C)_2Si=Si=Si(CH_3)_2$  (**4**) at the B3LYP/6-31+G(*d,p*) level<sup>10</sup> (Fig. 3) showed that the optimized structure **4[opt]** is characterized to be neither linear nor bent allenic but zwitterionic; the  $Si^1-Si^2-Si^3$  skeleton is remarkably bent with the bond angle of  $92.4^\circ$  and two  $C-Si-C$  planes are slightly twisted from each other (the dihedral angle  $C^2-Si^1-Si^2-Si^3$  is  $26.6^\circ$ ). The highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO, respectively) for **4[opt]** are a symmetric and an antisymmetric  $\pi$  orbital delocalized to the trisilaallyl  $\pi$  system, respectively. The reason for the discrepancy between the experimental structure for **1** and theoretical **4[opt]** may be ascribed to the severe steric hindrance between two bulky silacyclopentane rings caused when a zwitterionic structure similar to **4[opt]** is applied to **1**. To reveal the energy and electronic structure of bent allenic trisilaallene **1**, the molecular orbital calculations were carried out for **4[exp]**, where the coordinates of four carbon and three silicon atoms in **4** were fixed to those observed for **1**. As shown in Fig. 3, in **4[exp]**, twisted  $p\pi$  type orbitals at the terminal silicon atoms interact with both  $p\pi$ -type and in-plane orbitals at the central silicon atom, indicating the bent-allenic electronic structure in **4[exp]** and hence in **1**. Because **4[exp]** is only  $8.8 \text{ kcal mol}^{-1}$  higher in energy than **4[opt]**, the distortion from a zwitterionic to a bent-allenic structure will be feasible by the steric effects. In other words, the bulky terminal substituents in **1** not only serve as sterically protecting groups for  $Si=Si$  bonds but also adjust the electronic structure.

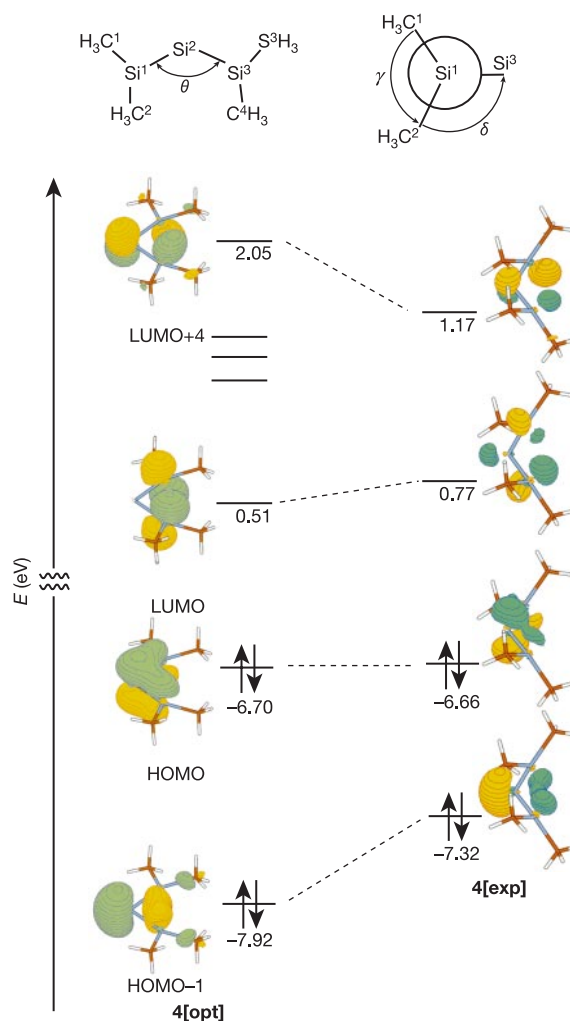
The unusual bonding nature of **1** in the solid state is also applicable to that in solution; the  $Si^1-Si^2-Si^3$  skeleton is bent and the  $Si^2$  atom rotates rapidly around the  $Si^1-Si^3$  axis. The  $^{29}Si$  NMR resonances found at 157 and 197 p.p.m. in benzene- $d_6$  are assigned to  $Si^2$  and  $Si^1$  ( $Si^3$ ) nuclei, respectively, being in good accord with the theoretical values calculated at the GIAO/6-311+G(2*df,p*)/B3LYP/6-31 + G(*d,p*) level for **4[exp]**; the theoretical  $^{29}Si$  resonances for  $Si^2$  and  $Si^1$  ( $Si^3$ ) nuclei for **4[exp]** are 144 and 221(212) p.p.m., respectively, whereas they are 39 and 155 p.p.m. for **4[opt]**. The fluxional nature of **1** in solution is evidenced by the highly symmetric  $^1H$ ,  $^{13}C$  and  $^{29}Si$  NMR spectra; 72  $^1H$ , 24  $^{13}C$ , and 8  $^{29}Si$  nuclei in 8 trimethylsilyl groups and 8  $^1H$  and 4  $^{13}C$  in two methylene groups are all equivalent. The  $^1H$  NMR spectrum for **1** was essentially independent of temperatures in toluene- $d_8$  and showed two singlets due to the methyl and methylene protons



**Figure 2** Newman projection illustrating the structure of the  $Si=Si=Si$  bonding in **1**. **a**, Newman projection along the  $Si^1-Si^2$  axis of **1**. **b**, Newman projection along the  $Si^3-Si^2$  axis. The Newman projections show clearly that the geometries around  $Si^1-Si^2$  and  $Si^3-Si^2$  double bonds are not planar but *trans*-bent.

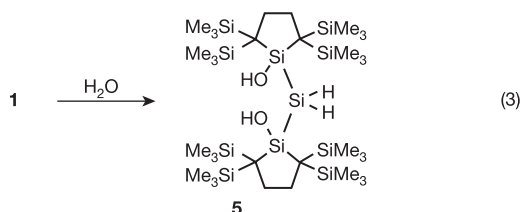
even at  $-80^\circ C$ , indicating low energy barriers for the interchange among the **A** to **D** isomers. The ultraviolet-visible (UV-vis) absorption maxima were found at 390 nm ( $\epsilon$  21300) and 584 nm ( $\epsilon$  700) in hexane; a UV-vis spectrum of **1** in hexane is given in the Supplementary Information. The remarkable red shift of the longest absorption band compared with that for tetramethyldisilene ( $\lambda_{max} = 350 \text{ nm}$ ) (ref. 11) as well as two-splitting band feature for **1** are consistent with significant conjugation between the two  $Si=Si$  double bonds as suggested by the molecular orbital calculations; the extent of the conjugation is even larger than that found for a spiropentasiladiene<sup>12</sup>.

The reaction of **1** with water gave the corresponding 1,3-dihydroxytrisilane **5** in almost quantitative yield as expected from the bent allenic structure (equation (3)) (see the Supplementary



**Figure 3** Schematic MO diagram for tetramethyltrisilaallene **4** depending on the skeletal deformation. *Ab initio* MO calculations were performed at the B3LYP/6-31+G(*d,p*) level. The structures of **4[opt]** (**4** with optimized structure) and **4[exp]** (**4** with the same skeletal structure as that found for **1A**) are characterized by the  $Si^1-Si^2-Si^3$  bent angle ( $\theta$ ),  $C^1-Si^1-Si^2-C^2$  dihedral angle ( $\gamma$ ) and  $C^2-Si^1-Si^2-Si^3$  dihedral angle ( $\delta$ ). A set of  $\theta$ ,  $\gamma$  and  $\delta$  in degrees is  $92.4$ ,  $162.9$  and  $26.6$  for **4[opt]** and  $136.5$ ,  $146.9$  and  $65.6$  for **4[exp]**. In **4[opt]**, the highest occupied and the lowest unoccupied molecular orbitals (HOMO and LUMO) are both  $\pi$ -type but HOMO-1 is in-plane  $n$ -type, indicating the zwitterionic nature of **4[opt]**. On the other hand, in **4[exp]**, the LUMO, LUMO+1, HOMO and HOMO-1 are all twisted bonding  $\pi$  and antibonding  $\pi^*$  orbitals, and hence, **4[exp]** is characterized as a bent-allenic molecule with the significant interaction between the two twisted  $Si=Si$  double bonds. Although we expected the two  $\pi$  as well as the two  $\pi^*$  orbitals to be degenerate at the linear-perpendicular geometry of **4**, no energy minimum was found at that geometry.

Information for the experimental details, spectral data and X-ray analysis for 5).



The bent-allenic and conjugated structure with facile intramolecular rotation of the Si<sup>1</sup> atom around the Si<sup>1</sup>–Si<sup>3</sup> axis found for the first stable trisilaallene is far from a reasonable extension of the bonding description for carbon allenes. Trisilaallene derivatives are expected to be useful synthetic reagents and building blocks for silicon functional materials in the future because of their unusual electronic structure. □

## Methods

### Synthesis of trisilaallene 1

When tetrachlorodisilane 3 (279 mg, 0.52 mmol) was treated with potassium graphite KC<sub>8</sub> (311 mg, 2.30 mmol) in THF (45 ml) at –40 °C for 24 h, the solution turned to dark green. Changing the solvent to hexane, elimination of insoluble materials by filtration, and then concentration of the solution to 1 ml produced crystals of trisilaallene 1 (117 mg, 0.15 mmol, 59%). Recrystallization from hexane afforded single crystals suitable for X-ray analysis.

### Spectral data for 1

Dark-green crystals with a melting point of 198–200 °C. <sup>1</sup>H NMR spectrum (C<sub>6</sub>D<sub>6</sub>, chemical shift δ in p.p.m.): 0.39 (singlet, 2H), 2.00 (singlet, 8H). <sup>13</sup>C NMR spectrum (C<sub>6</sub>D<sub>6</sub>, δ): 3.3 (SiMe<sub>3</sub>), 27.2 (C), 34.3 (CH<sub>2</sub>). <sup>29</sup>Si NMR spectrum (C<sub>6</sub>D<sub>6</sub>, δ): 1.6 (SiMe<sub>3</sub>), 157.0 (Si=Si=Si), 196.9 (Si=Si=Si). UV–vis spectrum (hexane): λ<sub>max</sub> = 390 nm (ε = 21300), λ<sub>max</sub> = 584 (ε = 700). Mass spectrometry (70 eV, electron impact ionization): m/z = 772 (molecular ion; relative intensity, 10%), m/z = 373 (41%), m/z = 73 (100%).

Received 21 October; accepted 17 December 2002; doi:10.1038/nature01380.

- Jutzi, P. Stable systems with a triple bond to silicon or its homologues: Another challenge. *Angew. Chem. Int. Edn* **39**, 3797–3800 (2000).
- Stender, M., Phillips, A. D., Wright, R. J. & Power, P. P. Synthesis and characterization of a digermanium analogue of an alkyne. *Angew. Chem. Int. Edn* **41**, 1785–1787 (2002).
- Phillips, A. D., Wright, R. J., Olmstead, M. M. & Power, P. P. Synthesis and characterization of 2,6-Dipp<sub>2</sub>-H<sub>3</sub>C<sub>6</sub>SnSnC<sub>6</sub>H<sub>5</sub>-2,6-Dipp<sub>2</sub> (Dipp = C<sub>6</sub>H<sub>3</sub>-2,6-Pr<sub>2</sub>): A tin analogue of an alkyne. *J. Am. Chem. Soc.* **124**, 5930–5931 (2002).
- Pu, L., Twamley, B. & Power, P. P. Synthesis and characterization of 2,6-Trip<sub>2</sub>-H<sub>3</sub>C<sub>6</sub>PbPbC<sub>6</sub>H<sub>5</sub>-2,6-Trip<sub>2</sub> (Trip = C<sub>6</sub>H<sub>2</sub>-2,4,5-Pr<sub>3</sub>): A stable heavier group 14 element analogue of an alkyne. *J. Am. Chem. Soc.* **122**, 3524–3525 (2000).
- Wiberg, N. et al. Tetrasupersilyl-tristannaallene and tristannacyclopentene (t-Bu<sub>3</sub>Si)<sub>4</sub>Sn<sub>3</sub>—Isomers with the shortest Sn = Sn double bonds to date. *Eur. J. Inorg. Chem.* 1211–1218 (1999).
- Wiberg, N. et al. Synthesis, structure and dehalogenation of the disilene RClSi = SiClR [R = (t-Bu<sub>3</sub>Si)<sub>2</sub>MeSi]. *Eur. J. Inorg. Chem.* 1066–1070 (2002).
- Escudie, J., Ranaivonjatovo, H. & Rigon, L. Heavy allenes and cumulenes E = C = E' and E = C = C = E' (E = P, As, Si, Ge, Sn; E' = C, N, P, As, O, S). *Chem. Rev.* **100**, 3639–3696 (2000).
- Kira, M., Ishida, S., Iwamoto, T. & Kabuto, C. The first isolable dialkylsilylene. *J. Am. Chem. Soc.* **121**, 9722–9723 (1999).
- Harada, J. & Ogawa, K. Invisible but common motion in organic crystals: A pedal motion in stilbenes and azobenzenes. *J. Am. Chem. Soc.* **123**, 10884–10888 (2001).
- Frisch, M. J. et al. *Gaussian 98* Revision A.7 (Gaussian, Inc., Pittsburgh, PA, 1998).
- Sekiguchi, A., Maruki, I., Ebata, K., Kabuto, C. & Sakurai, H. High-pressure synthesis, structure, and novel photochemical reactions of 7,7,8,8-tetramethyl-7,8-disilabicyclo[2.2.2]octa-2,5-diene. *J. Chem. Soc. Chem. Commun.* 341–343 (1991).
- Iwamoto, T., Tamura, M., Kabuto, C. & Kira, M. A stable bicyclic compound with two Si = Si double bonds. *Science* **200**, 504–506 (2000).

Supplementary Information accompanies the paper on Nature's website (http://www.nature.com/nature).

**Acknowledgements** This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to M.K. (e-mail: mkira@si.chem.tohoku.ac.jp).

## Coral record of increased sediment flux to the inner Great Barrier Reef since European settlement

Malcolm McCulloch\*, Stewart Fallon\*, Timothy Wyndham\*, Erica Hendy\*, Janice Lough† & David Barnes‡

\* Research School of Earth Sciences, Australian National University, Canberra 0200, Australia

‡ Australian Institute of Marine Science, Townsville, Queensland 4810, Australia

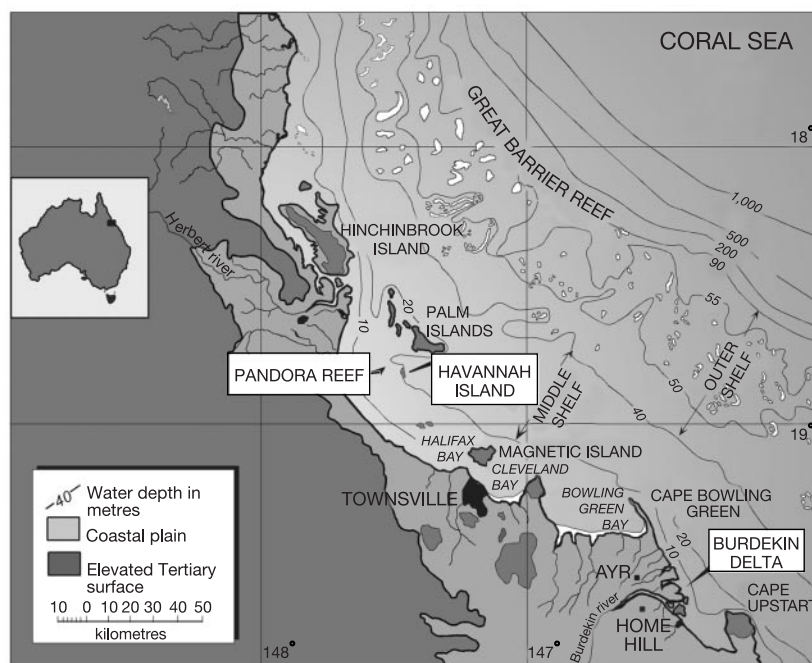
The effect of European settlement on water quality in the Great Barrier Reef of Australia is a long-standing and controversial issue<sup>1–6</sup>. Erosion and sediment transport in river catchments in this region have increased substantially since European settlement<sup>6–10</sup>, but the magnitude of these changes remains uncertain<sup>1–10</sup>. Here we report analyses of Ba/Ca ratios in long-lived *Porites* coral from Havannah Reef—a site on the inner Great Barrier Reef that is influenced by flood plumes from the Burdekin river—to establish a record of sediment fluxes from about 1750 to 1998. We find that, in the early part of the record, suspended sediment from river floods reached the inner reef area only occasionally, whereas after about 1870—following the beginning of European settlement—a five- to tenfold increase in the delivery of sediments is recorded with the highest fluxes occurring during the drought-breaking floods. We conclude that, since European settlement, land-use practices such as clearing and overstocking have led to major degradation of the semi-arid river catchments, resulting in substantially increased sediment loads entering the inner Great Barrier Reef.

Since European settlement, large-scale modification of the river catchments has occurred from grazing, agriculture, mining and associated activities such as land clearing. Their impact on the marine environment remains highly uncertain<sup>1–6,11,12</sup>. With few exceptions<sup>13</sup>, it has been difficult conclusively to demonstrate a link between enhanced levels of terrestrial runoff and large-scale reductions in coral reef cover. This is due not only to the complex sequence of events that may ultimately lead to a shift from coral- to algae-dominated communities<sup>11,12</sup>, but also because it has been difficult to quantify long-term changes in water quality regimes. Here we describe a new approach based on the application of *in situ* geochemical tracers in corals<sup>14–16</sup> that has the advantage of providing a direct quantitative measure of the long-term changes in sediment fluxes that are actually being delivered to the Great Barrier Reef (GBR).

Long-lived (300 to 400 years old) *Porites* corals from the Havannah and Pandora reefs, located on the inner GBR of northern Queensland, experience episodic discharges of freshwater flood plumes from the Burdekin river (Fig. 1). The Burdekin river is the largest single contributor of suspended sediment to the inner GBR lagoon, and often delivers more than 10<sup>7</sup> tonnes of suspended sediment during single runoff events<sup>6–10</sup>. During periods of flood, freshwater discharges form buoyant, low-salinity flood plumes that generally move northwards along the GBR coastline<sup>17</sup>. Occasionally flood plumes may also reach the mid-shelf region of the GBR and thus directly affect the more 'pristine' regions of the GBR<sup>17,18</sup>. Flood events are manifested in the coral as narrow fluorescent<sup>19</sup> or luminescent<sup>20</sup> bands, thought to be due to the incorporation of either humic acids or low-density zones in the coral skeleton. Here we show how Ba/Ca ratios in coral cores can provide high-fidelity records of suspended sediment loads entering the GBR. Barium is desorbed from fine-grained suspended particles (clays) in the low-

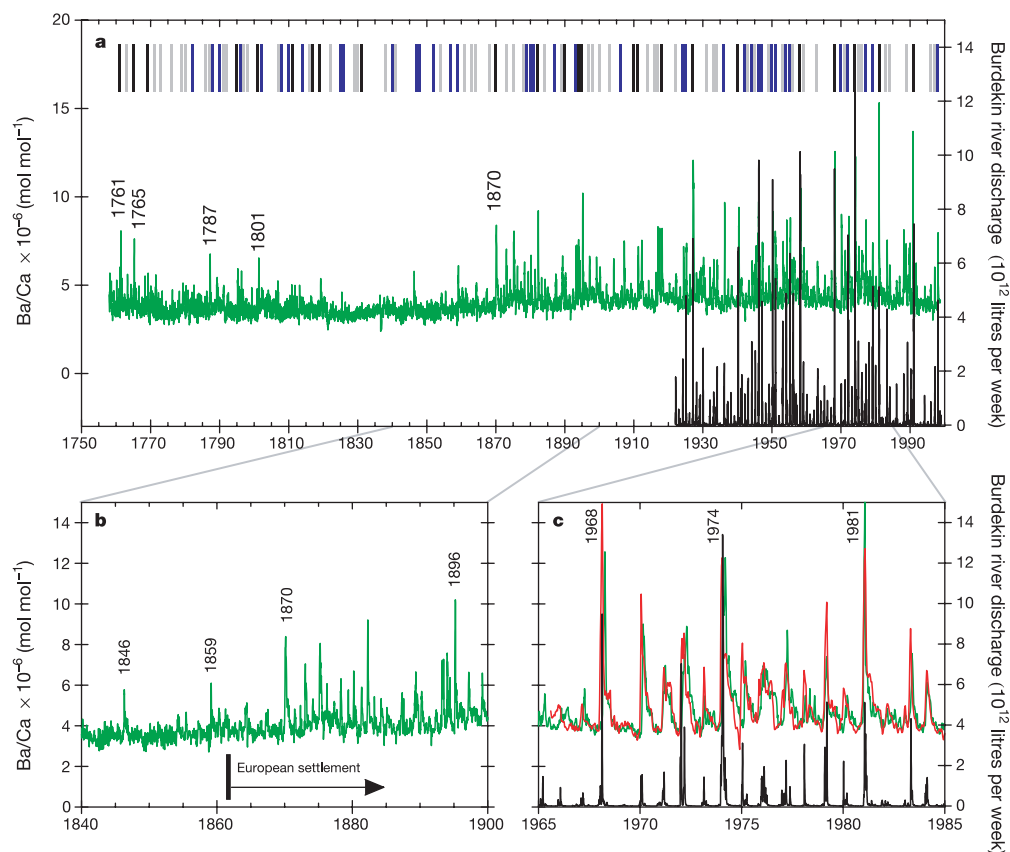
† Present address: Lawrence Livermore National Laboratory, California, USA.





**Figure 1** Map showing the location of the central Great Barrier Reef (after Belperio<sup>7</sup>). During flood events, low-salinity plumes are discharged from the Burdekin river and advected in a northwards direction within the inner-mid-shelf region of the GBR<sup>17</sup>. Coral

cores were analysed from the inshore Havannah and nearby Pandora reefs, which are in the pathway of Burdekin river flood plumes.



**Figure 2** The coral Ba/Ca record of suspended sediment into the GBR by the Burdekin river over approximately the past 250 years. Barium provides a proxy for suspended sediments as it is desorbed from flood plume sediments<sup>21</sup> and quantitatively partitioned into the corals, calcium carbonate skeleton<sup>22</sup>. **a**, Coral record from Havannah Reef (green line) for the period from 1760 to 1998 with Ba/Ca peaks proportional to the sediment flux delivered by Burdekin flood plume events. The frequency and intensity of flood events is indicated by luminescent bands in the coral skeleton<sup>19,20,28</sup>, represented here as major

(black bar), average (blue bar) and small (grey bar) discharge events. The record of discharge for the Burdekin river, available from 1921, is also shown (black line). **b**, Coral record for the period from 1840 to 1990 showing the large increase in Ba/Ca that commences in 1870, the first major flood following European settlement. **c**, Coral records from Havannah (green line) and Pandora reefs (red line) for the period 1965 to 1985 show excellent agreement and a good correlation with the Burdekin river weekly discharge (black line).

salinity region (0–5 p.p.t.) of the estuarine mixing zone<sup>21</sup>. Thereafter Ba behaves as an essentially conservative dissolved tracer, being advected with the flood plume and partitioned into the coral carbonate skeleton in proportion to the ambient seawater concentration<sup>22</sup>. Ba/Ca ratios in corals<sup>23</sup> therefore provide a long-term record of changes in suspended sediment loads and hence in the nutrients that are entering the GBR.

Using the relatively new technique of laser ablation inductively coupled plasmas mass spectrometry (ICP-MS) specifically adapted for coral studies<sup>14–16</sup>, it is now possible to investigate the detailed geochemical characteristics of entire coral cores, at approximately weekly resolution. A 5.3-m-long Havannah coral core was analysed for the period from about 1750 to 1985, supplemented by shorter cores collected in 1998 from Havannah and the nearby Pandora Reef. The Ba/Ca systematics in the coral core reveal two distinctive patterns (Fig. 2). Prior to European settlement there is surprisingly little evidence for flood-plume related activity from the coral Ba/Ca ratios. The only significant peaks that are present correspond to the major floods of 1761, 1765, 1787 and 1801 (Fig. 2a). From the early 1800s to 1860, which includes major flood events in the years 1811, 1817, 1819 and 1831, the coral does not exhibit any significant Ba peaks, despite the presence of major luminescent bands (Fig. 2a).

Soon after European settlement, which commenced in the Burdekin catchment in 1862 (ref. 24), there is a large change in the behaviour of Ba. The 1870 flood band (Fig. 2b), the first major flood following European settlement, has a large Ba/Ca peak, indicative of a significant increase in suspended load being delivered to the inner GBR. Large peaks also occurred soon after in 1875, followed by major peaks accompanying the floods in 1882 and 1896. The 1896 flood, from cyclone *Sigma*, produced a large Ba/Ca peak ( $> 10 \times 10^{-6}$ ), despite a flood level of only 15.8 m, compared to the first recorded flood height of 20.9 m for the 1870 flood.

Another important feature of the Havannah coral record is the significant increase in the Ba/Ca baseline that also occurred from around 1870 onwards (Fig. 2a, b). Baseline ratios increase from the pre-European range of  $(3\text{--}4) \times 10^{-6}$ , to a present-day range of  $(4\text{--}5) \times 10^{-6}$ , an increase of about 30%. The sustained increase in the Ba/Ca baseline that commences after the 1870 flood is interpreted as indicating a finite but significant residence time of at least several years, for an enhanced flux of terrestrially derived Ba. Fluctuations in the Ba/Ca baseline and the apparent persistence after some large flood events (indicated by shoulders on the Ba peaks, Fig. 2c), suggests that processes such as biological recycling<sup>21</sup>, continuing Ba supply from estuaries<sup>16</sup>, and possibly groundwater inputs<sup>25</sup> to the GBR may also make second-order contributions to the inshore Ba budget. Upwelling of Ba-rich waters may also contribute to fluctuations in the Ba/Ca baseline<sup>22,26</sup>, but these effects are relatively minor and mainly restricted to the outer reef. Together, these results indicate that within one to two decades after the arrival of European settlers in northern Queensland, there were already massive impacts on the river catchments that were being transmitted to the waters of the inner GBR. This is mainly attributed to the rapid expansion in numbers of (initially) sheep and, most importantly, cattle<sup>24</sup> (Table 1) and the resultant erosion that introduced hoofed animals would have had on the highly fragile riverbanks of the Burdekin.

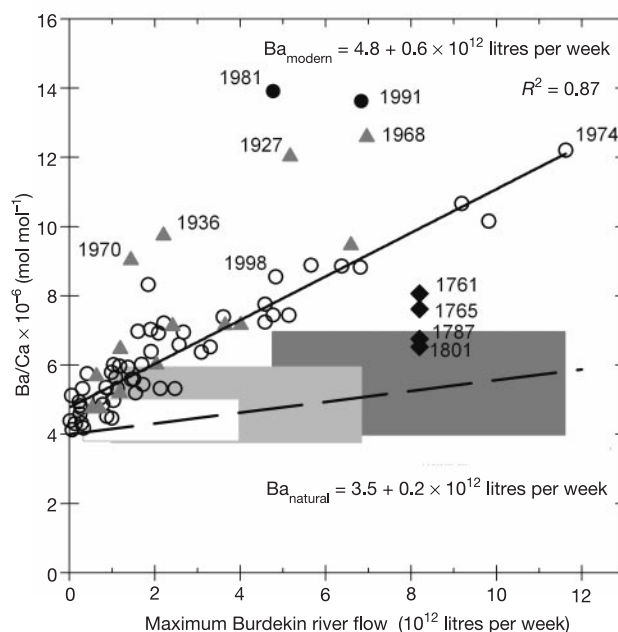
The relationship between Ba/Ca in corals and the Burdekin river discharge is shown in Fig. 3, where Ba/Ca peak heights are plotted versus the maximum weekly river discharge using records that

commence in 1921. Weekly discharge is used, as this is the time-frame for transport of flood plumes from the mouth of the Burdekin river to the Havannah reef site (Fig. 1). There is an approximately linear relationship between the maximum weekly discharge of the Burdekin river and the maximum coral Ba/Ca ratio ( $r^2 = 0.68$ ), which is greatly improved if floods that follow periods of drought (peak Burdekin flow  $< 0.7 \times 10^{12}$  litres per week) are excluded (Fig. 3). The relationship for non-drought modern (1921 to 1998) floods is given by:

$$\text{Ba}_{\text{modern}} = 4.8 + 0.6 \times 10^{12} \text{ litres per week} \quad (1)$$

where  $\text{Ba} = \text{Ba/Ca} \times 10^{-6}$  (atomic ratios) and the intercept is  $4.8 \times 10^{-6}$ . If the floods of 1981 and 1991 are excluded (see later discussion), then a highly coherent array ( $r^2 = 0.87$ ) is obtained. The observation that drought-breaking floods have much greater suspended sediment loads is not surprising considering the loss of groundcover and hence the enhanced erosion that occurs during periods of drought. Prominent examples are the runoff events of 1927, 1936, 1968, 1970 and 1988, which were all preceded by droughts. The higher Ba/Ca ratios in drought-breaking floods indicates that suspended sediment loads are approximately doubled, relative to the  $\text{Ba}_{\text{modern}}$  reference line. The 1981 flood is highly anomalous, possibly reflecting expansion of landclearing<sup>6,27</sup> and the large increase in cattle stocking in the Burdekin catchment that occurred during the 1970s, with the introduction of more drought-resistant cattle breeds from India, such as *Bos indicus*<sup>27</sup>. The very large flood event of 1991 is also unusual, as it is a double-pulse flood event, and hence estimates of the weekly discharge maximum are more sensitive to the flood hydrograph.

The natural or pre-European influence of the Burdekin river on



**Figure 3** Plot of the maximum Ba/Ca flood peak height versus the maximum weekly Burdekin river discharge for flood events from 1921 to the present. Triangles correspond to floods that follow periods of drought (that is, peak Burdekin flow  $< 0.7 \times 10^{12}$  litres per week). If the events of 1981 and 1991 are also excluded (see text) then there is an excellent correlation ( $r^2 = 0.87$ ) between Ba/Ca (that is, sediment flux) and maximum weekly discharge for flood events. The magnitude of pre-European floods is estimated using both luminescent bands<sup>28</sup> and Sr/Ca- $\delta^{18}\text{O}$  systematics<sup>29</sup> (see Supplementary Information for details). Luminescent bands for pre-European floods are shown for major floods (dark grey), average (light grey) and small discharge events (open rectangle). Solid diamonds show the magnitude of the largest pre-European floods.

**Table 1** Cattle and sheep numbers in the Townsville region<sup>24,27</sup>

|        | 1862 | 1868    | 1872   | 1889    | 1990       |
|--------|------|---------|--------|---------|------------|
| Sheep  | 0    | 176,956 | 1,284  | 2,983   | -          |
| Cattle | 0    | 31,098  | 53,457 | 250,912 | >1,000,000 |

the inner GBR can be determined using the same approach and is given by:

$$Ba_{\text{natural}} = 3.5 + 0.2 \times 10^{12} \text{ litres per week} \quad (2)$$

Where pre-historic Burdekin river flows (that is, before 1860) have been determined using the relative intensity of luminescent bands<sup>19,20,28</sup> as shown in Fig. 2, together with more limited  $\delta^{18}\text{O}$ -Sr/Ca systematics<sup>29</sup> (see Supplementary Information). The increase in both the slope and baseline of the Ba/Ca versus river-discharge relationships—equations (1) and (2)—indicates a five- to tenfold increase in suspended sediment load following European settlement of the Burdekin river catchment. The most pronounced deviation from the pre-European reference line are the floods of 1761, 1765, 1787 and 1801, all of which occur after periods of drought. This shows that even in its 'natural' state the Burdekin river catchment was subjected to enhanced erosion due to drought, albeit at a reduced level. It is also noted that from 1850 to 1870 a shift in  $\delta^{18}\text{O}$  in coral cores has been observed<sup>30</sup>, indicating freshening of the waters of the inner GBR. Although this has been attributed<sup>30</sup> to a regional scale phenomenon marking the end of the Little Ice Age, there remains the possibility that land-use changes since European settlement have also increased freshwater discharges into the inner GBR. Regardless, it is clear that climate change (especially droughts), combined with the adoption of European-style land use has had a severe impact on river catchments and hence on the near-shore coral reefs of the GBR.

The impact of higher loads of sediment and nutrients from river runoff in the GBR is difficult to quantify, but past experience<sup>11–13</sup> suggests it should cause concern. It is clear, for example, that dissolved components such as inorganic P, organic P and N, nitrite and nitrate, are enriched in flood plumes<sup>2,10</sup> and often result in higher levels of biologic activity such as phytoplankton blooms<sup>2</sup>. Fortunately for the GBR, these effects are probably mainly limited to the duration of flood plumes, that is, from weeks to months. Arguably<sup>1–6</sup>, of most concern are the direct (for example, increased turbidity) and indirect (for example, desorption of P from anoxic sediments) impacts of increased sediment loads in the GBR. Burdekin river flood plumes supply an average of  $10^{10} \text{ m}^3$  of water and  $10^7$  tonnes per year of suspended sediment to the inner GBR<sup>6–10</sup>. Suspended sediment concentrations measured close to the river mouth are typically  $500 \text{ mg l}^{-1}$  to  $1,500 \text{ mg l}^{-1}$  (refs 2, 10), consistent with the range expected from estimates from erosion and soil loss in the catchment<sup>6,8,9</sup>. More distal from the river mouth ( $>10 \text{ km}$ ), suspended sediment concentrations generally decrease stepwise to levels of approximately  $10\text{--}20 \text{ mg l}^{-1}$ , concomitant with increasing salinity<sup>2</sup>. This indicates that only about 5% to 10% of the sediment remains suspended in flood plumes. Nevertheless, flood plumes from the Burdekin river have the potential to disperse large amounts of sediment ( $10^5$  to  $10^6$  tonnes) over a significant region (for example,  $500 \text{ km} \times 20 \text{ km}$ ) of the inner and mid-GBR. The fate of the more than  $10^9$  tonnes of additional sediment deposited at the river mouth since European settlement commenced has not been considered in this budget, but depending on the tide- or wind-driven lateral transport processes is also likely to have significant impacts.

Fine-grained suspended sediments are likely to have the most deleterious effects on coral reefs. These sediments can be readily transported and dispersed throughout the inner to mid-shelf regions of the GBR, and are more chemically reactive, being the major source of terrestrial P and other important nutrient elements. Reducing sediment fluxes to coral reefs must be a high priority if corals are to survive the damaging combination of direct anthropogenic impact and rapid climate change. □

Received 6 June; accepted 10 December 2002; doi:10.1038/nature01361.

1. Bell, P. R. F. Status of eutrophication in the Great Barrier Reef Lagoon. *Mar. Pollut. Bull.* **22**, 89–93 (1991).

- Devlin, M., Waterhouse, J., Taylor, J. & Brodie, J. Flood plumes in the Great Barrier Reef: spatial and temporal patterns in composition and distribution. *Great Barrier Reef Mar. Park Auth.* **68** (2001).
- Haynes, D. & Michalek-Wegner, K. Water quality in the Great Barrier World Heritage Area: past perspectives, current issues and new research directions. *Mar. Pollut. Bull.* **41**, 428–434 (2000).
- Larcombe, P. & Woolfe, K. Increased sediment supply to the Great Barrier Reef will not increase sediment accumulation at most coral reefs. *Coral Reefs* **18**, 163–169 (1999).
- McCook, L. J. Macroalgae, nutrients and phase shifts on coral reefs: scientific issues and management consequences for the Great Barrier Reef. *Coral Reefs* **18**, 357–367 (1999).
- Neil, D. T., Orpin, A. R., Ridd, P. V. & Bofu, Y. Sediment yield and impacts from river catchments to the Great Barrier Reef lagoon. *Mar. Freshwat. Res.* **53**, 733–752 (2002).
- Belperio, A. P. Terrigenous sedimentation in the central Great Barrier Reef Lagoon: a model from the Burdekin region. *Bur. Miner. Resour. J. Aust. Geol. Geophys.* **8**, 179–190 (1983).
- Moss, A. J., Raymond, G. E., Reilly, N. & Best, E. K. *A Preliminary Assessment of Sediment and Nutrient Exports from Queensland Coastal Catchments* (Queensland Department of Environment and Heritage, Brisbane, 1993).
- Prosser, I. P. et al. Large-scale patterns of erosion and sediment transport in river networks, with examples from Australia. *Mar. Freshwat. Res.* **52**, 81–99 (2001).
- Furnas, M. J. & Mitchell, A. W. in *Oceanographic Processes of Coral Reefs: Physical and Biological Links in the Great Barrier Reef* (ed. Wolanski, E.) 37–52 (CRC Press, Boca Raton, Florida, 2001).
- Done, T. J. Phase shifts in coral reefs communities and their ecological significance. *Hydrobiologia* **247**, 121–155 (1992).
- Hughes, T. P. Catastrophes, phases shifts, and large-scale degradation of a Caribbean coral reef. *Science* **265**, 1547–1551 (1994).
- Smith, S. V., Kimmerer, W. J., Laws, E. A., Brock, R. E. & Walsh, T. W. Kaneohe Bay sewage diversion experiment: perspectives on ecosystem responses to nutritional perturbation. *Pacif. Sci.* **35**, 279–385 (1981).
- Sinclair, D., Kinsley, L. P. & McCulloch, M. T. High resolution analysis of trace elements in corals by laser ablation ICP-MS. *Geochim. Cosmochim. Acta* **62**, 1889–1901 (1997).
- Fallon, S. J., White, J. C. & McCulloch, M. T. Porites corals as recorders of mining and environmental impacts: Misima Island, Papua New Guinea. *Geochim. Cosmochim. Acta* **66**, 45–62 (2002).
- Alibert, C. et al. Source of trace element variability in Great Barrier Reef corals affected by the Burdekin flood plumes. *Geochim. Cosmochim. Acta* (in the press).
- King, W. E., McAllister, F., Woolfe, K., Done, T. & Spagnol, S. in *Oceanographic Processes of Coral Reefs: Physical and Biological Links in the Great Barrier Reef* (ed. Wolanski, E.) 145–159 (CRC Press, Boca Raton, Florida, 2001).
- McCook, L. J., Wolanski, E. & Spagnol, S. in *Oceanographic Processes of Coral Reefs: Physical and Biological Links in the Great Barrier Reef* (ed. Wolanski, E.) 113–126 (CRC Press, Boca Raton, Florida, 2001).
- Isdale, P. J. Fluorescent bands in massive corals record centuries of coastal rainfall. *Nature* **310**, 578–579 (1984).
- Barnes, D. J. & Taylor, R. B. On the nature and causes of luminescent lines and bands in coral skeletons. *Coral Reefs* **19**, 221–230 (2001).
- Li, Y. H. & Chan, L. H. Desorption of Ba and <sup>226</sup>Ra from river-borne sediments in the Hudson estuary. *Earth Planet. Sci. Lett.* **43**, 343–350 (1979).
- Lea, D. W., Shen, G. T. & Boyle, E. A. Coralline barium records temporal variability in Equatorial Pacific upwelling. *Nature* **340**, 373–376 (1989).
- Shen, G. T. & Sanford, C. L. in *Global Consequences of the 1982–83 El Niño/Southern Oscillation* (ed. Glynn, P. W.) 255–284 (Elsevier, New York, 1990).
- Holmes, J. M. *Australia's Open North* (Angus & Robertson, Sydney, 1963).
- Moore, W. S. The subterranean estuary: a reaction zone of ground water and sea water. *Mar. Chem.* **65**, 111–125 (1999).
- Tudhope, A. W., Lea, D. W., Shimmield, G. B., Chilcott, C. P. & Head, S. Monsoonal climate and Arabian Sea Coastal upwelling recorded in massive corals from Southern Oman. *Paleosols* **11**, 347–361 (1996).
- Ash, A. J., Mclvor, J. G., Mott, J. J. & Andre, M. H. Building grass castles: integrating ecology and management of Australia's tropical tallgrass rangelands. *Rangel. J.* **19**, 123–144 (1997).
- Lough, J. M., Barnes, D. J. & McAllister, F. Luminescent lines in corals from the Great Barrier Reef provide spatial and temporal records of reefs affected by land runoff. *Holocene Res.* (in the press).
- McCulloch, M. T., Gagan, M. K., Mortimer, G. E., Chivas, A. R. & Isdale, P. J. A high resolution Sr/Ca and  $\delta^{18}\text{O}$  coral record from the Great Barrier Reef, Australia and the 1982–1983 El Niño. *Geochim. Cosmochim. Acta* **58**, 2747–2754 (1994).
- Hendy, E. et al. Abrupt decrease in tropical Pacific Sea surface salinity at end of little ice age. *Science* **295**, 1511–1514 (2002).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com/nature>).

**Acknowledgements** This study has greatly benefited from developments in laser ablation instrumentation undertaken by S. Eggins, M. Shelley and L. Kinsley. Appreciation is extended to D. Sinclair whose initial work stimulated this research. We thank M. Devlin, D. Haynes, K. Fabricius, J. Brodie and J. Marshall for assistance in undertaking marine fieldwork and discussions on terrestrial impacts in the Great Barrier Reef. We thank P. Isdale for collecting the Havannah coral core. We are also grateful to A. Chivas for constructive comments on the manuscript. Data can be accessed at the World Data Centre for Paleoclimatology at [www.ngdc.noaa.gov/paleo](http://www.ngdc.noaa.gov/paleo). This research is supported by a grant from the ARC.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to M.M. (e-mail: [Malcolm.McCulloch@anu.edu.au](mailto:Malcolm.McCulloch@anu.edu.au)).



# A discontinuity in mantle composition beneath the southwest Indian ridge

Christine M. Meyzen<sup>\*†</sup>, Michael J. Toplis<sup>\*</sup>, Eric Humler<sup>‡</sup>, John N. Ludden<sup>\*</sup> & Catherine Mével<sup>‡</sup>

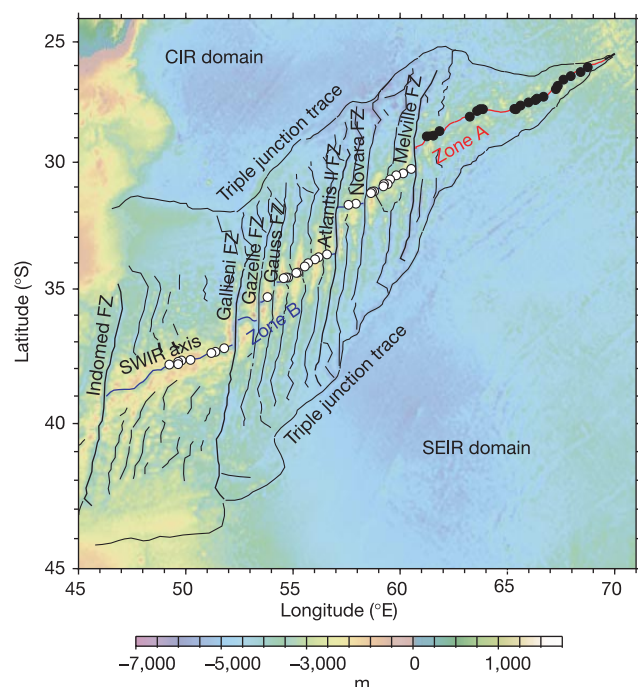
<sup>\*</sup> Centre de Recherches Pétrographiques et Géochimiques (CRPG), UPR 2300, 15 rue Notre-Dame-des-Pauvres, BP 20, 54501 Vandoeuvre-Les-Nancy, France  
<sup>‡</sup> Laboratoire de Géosciences Marines, Institut de Physique du Globe (IPG), 4 place Jussieu, 75251 Paris cedex 05, France  
<sup>†</sup> Present address: Danish Lithosphere Centre, Øster Voldgade 10, 1350 Copenhagen, Denmark

The composition of mid-ocean-ridge basalt is known to correlate with attributes such as ridge topography<sup>1,2</sup> and seismic velocity in the underlying mantle<sup>3</sup>, and these correlations have been interpreted to reflect variations in the average extent and mean pressures of melting during mantle upwelling. In this respect, the eastern extremity of the southwest Indian ridge is of special interest, as its mean depth of 4.7 km (ref. 4), high upper-mantle seismic wave velocities<sup>5</sup> and thin oceanic crust of 4–5 km (ref. 6) suggest the presence of unusually cold mantle beneath the region. Here we show that basaltic glasses dredged in this zone, when compared to other sections of the global mid-ocean-ridge system, have higher Na<sub>8,0</sub>, Sr and Al<sub>2</sub>O<sub>3</sub> compositions, very low CaO/Al<sub>2</sub>O<sub>3</sub> ratios relative to TiO<sub>2</sub> and depleted heavy rare-earth element distributions. This signature cannot simply be ascribed to low-degree melting of a typical mid-ocean-ridge source mantle, as different geochemical indicators of the extent of melting<sup>1</sup> are mutually inconsistent. Instead, we propose that the mantle

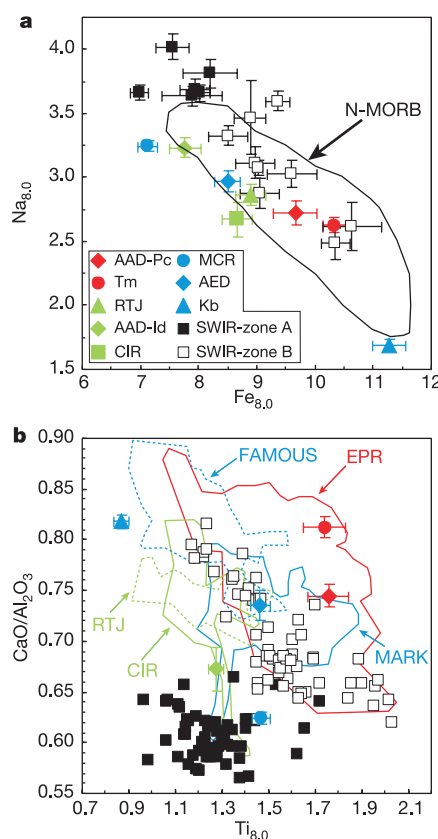
beneath ~1,000 km of the southwest Indian ridge axis has a complex history involving extensive earlier melting events and interaction with partial melts of a more fertile source.

This study reports the major- and trace-element compositions of basaltic glasses dredged along the southwest Indian ridge (SWIR) between 49° and 69° E during the EDUL cruise (Fig. 1, Supplementary Table 1). Although this entire section of ridge axis is characterized by extremely low spreading rate (full spreading rate, 11–15 mm yr<sup>-1</sup>)<sup>7</sup>, it may be divided (on the basis of geophysical characteristics) into two zones separated by the Melville fracture zone (MFZ) at 60° 45' E. To the east of the MFZ (zone A, Fig. 1) the ridge displays the extreme geophysical characteristics described above, while to the west of the MFZ (zone B, Fig. 1), the axial segmentation pattern is smoother and mean axial depth ranges from 4,330 to 3,090 m going from east to west.

Sample compositions have Mg number (Mg# = 100Mg/(Mg + Fe)) in the range 0.49–0.66, although most of the compositions located to the east of the MFZ are relatively undifferentiated (Mg# > 0.58). When compared at similar MgO contents, the major-element chemistry of basaltic glasses from zone A is systematically distinct from all previously described mid-ocean-ridge basalt (MORB; Supplementary Table 1). For example, these lavas



**Figure 1** Topographic map of the southwest Indian ridge (SWIR), showing sample locations used in this study (circles). (Map is from ref. 21.) The SWIR is propagating at about twice its spreading rate<sup>22</sup> at the Rodriguez triple junction, where the central Indian ridge (CIR) joins also the southeast Indian ridge (SEIR). FZ, fracture zone.



**Figure 2** Characteristics of the major-element chemistry of SWIR lavas compared to those from other mid-ocean ridges. **a**, Na<sub>8,0</sub> versus Fe<sub>8,0</sub>, and **b**, CaO/Al<sub>2</sub>O<sub>3</sub> versus Ti<sub>8,0</sub>. The subscript 8.0 indicates that oxides have been corrected to MgO = 8% (refs 1, 2). Na<sub>8,0</sub> and Fe<sub>8,0</sub> have been averaged over mean ridge length of 60 ± 30 km. Outlined fields are for normal MORB (N-MORB)<sup>2</sup>, East Pacific Rise (EPR), MARK, FAMOUS areas, central Indian ridge (CIR) and Rodriguez triple junction (RTJ). Also shown are averages from the Mid-Cayman rise (MCR), Indian and Pacific sides of the Australian–Antarctic discordance (AAD-Id and AAD-Pc respectively), Atlantic equatorial discordance (AED), Tamayo (Tm) region of the EPR, RTJ, CIR and Kolbeinsey ridge (Kb). Only data from the two deepest ridge segments of the AED are used. Error bars are 2σ. See Supplementary Information for references used.

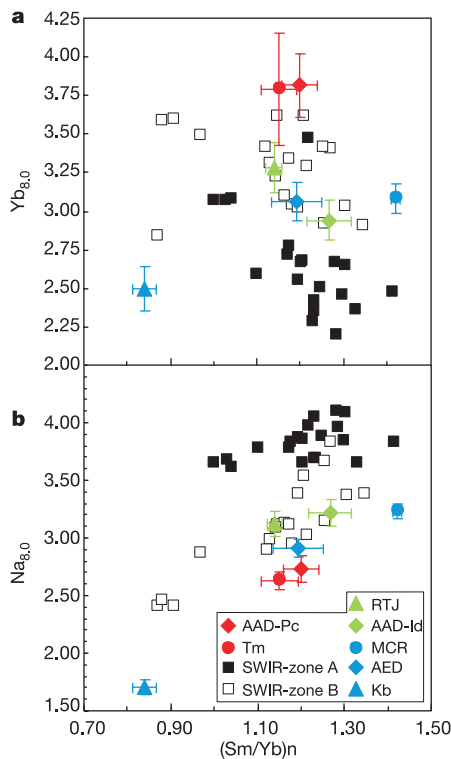
define the highest reported values of  $\text{Na}_{8.0}$  at low values of  $\text{Fe}_{8.0}$  in the global array<sup>1</sup> (Fig. 2a). In contrast, basalts from zone B fall in the global array of  $\text{Na}_{8.0}$  versus  $\text{Fe}_{8.0}$ . Here and elsewhere the subscript 8.0 refers to values corrected for low-pressure fractionation to a common MgO content of 8wt%, as described by ref. 1.

Basalts from zone A are also distinguished by their consistently high  $\text{Al}_2\text{O}_3$  (16.1–18.3 wt%) and low  $\text{Ti}_{8.0}$ , as illustrated by the co-variation of  $\text{CaO}/\text{Al}_2\text{O}_3$  and  $\text{Ti}_{8.0}$  (Fig. 2b). Even if rare lavas with low  $\text{CaO}/\text{Al}_2\text{O}_3$  and low  $\text{Ti}_{8.0}$  have been previously identified in the Atlantic equatorial discordance<sup>8</sup>, the MARK region<sup>9</sup> and the central Indian ridge<sup>10</sup>, it is clear that zone A lavas systematically define a unique compositional field for MORB glasses. The mid-Cayman rise<sup>11,12</sup> and, to a lesser extent, the Indian side of the Australian–Antarctic discordance<sup>13–15</sup>, two ‘cold sections’ of the global mid-ocean ridge system, also display mild offsets to low  $\text{CaO}/\text{Al}_2\text{O}_3$  relative to  $\text{Ti}_{8.0}$ , as do the fields defined by the Rodriguez triple junction and the central Indian ridge. In contrast, MORB from zone B falls within the compositional array defined by North Atlantic lavas.

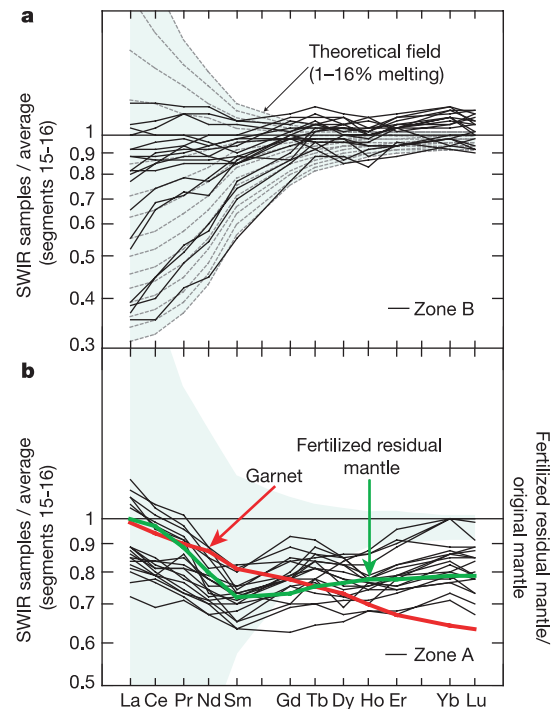
In terms of trace-element geochemistry, zone B glasses overlap trends defined by basalts from other mid-ocean spreading centres (see, for example, Fig. 3a). This is also true for highly incompatible elements such as  $\text{Ce}_{8.0}$  in basalts of zone A. However, when compared as a function of  $(\text{Sm}/\text{Yb})_n$  (where the subscript n means chondrite normalized values), these lavas are found to be progressively depleted in trace elements of increasing compatibility, such that they have uniformly lower heavy rare-earth element (HREE) concentrations relative to basalts from zone B and to the global array (Fig. 3a). This offset is also a feature of the Indian side of

the Australian–Antarctic discordance. Furthermore, consideration of  $\text{Na}_{8.0}$  as a function of  $(\text{Sm}/\text{Yb})_n$  demonstrates that zone A lavas are highly enriched in sodium compared to all other ridges at a given value of  $(\text{Sm}/\text{Yb})_n$  (Fig. 3b). The same is true of Sr and  $\text{Al}_2\text{O}_3$ .

Although the co-variation of  $\text{Na}_{8.0}$  and  $\text{Fe}_{8.0}$  (Fig. 2a) could justifiably lead to the inference that zone A basalts may be explained by extremely low extents of mantle melting at low mean pressure, consideration of all major and trace elements shows that this simple scheme is untenable. For example, during partial melting of a clinopyroxene-bearing source the  $(\text{Sm}/\text{Yb})_n$  of liquids should also be an indicator of the extent of partial melting<sup>1</sup>. However, although basalts from zone B show the same correlation between  $\text{Na}_{8.0}$  and  $(\text{Sm}/\text{Yb})_n$  as that defined by other spreading centres, this is clearly not the case for zone A lavas (Fig. 3b). Furthermore, as partial melting of a clinopyroxene-bearing source proceeds, the  $\text{CaO}/\text{Al}_2\text{O}_3$  of liquids increases while  $\text{TiO}_2$  decreases, leading to the co-variation observed for zone B lavas and the vast majority of lavas from the global array (Fig. 2b). Lavas from zone A are obviously inconsistent with this general trend, suggesting that a lower-than-normal degree of partial melting of a typical mantle source is not responsible for their unique geochemistry.



**Figure 3** Variations of major- and trace-element chemistry of SWIR lavas compared to those from other mid-ocean ridges.  $\text{Yb}_{8.0}$  (a) and  $\text{Na}_{8.0}$  (b) versus  $(\text{Sm}/\text{Yb})_n$ . Symbols as in Fig. 2. To avoid effects of low-pressure fractional crystallization, REE contents have been corrected to 8 wt% MgO. This correction is based on the calculation of  $\text{Ce}_{8.0}$  from ref. 23, and assuming that all REE have a partition coefficient of zero during low-pressure fractionation.



**Figure 4** Rare-earth element systematics of glasses from zones B and A. **a**, Zone B; **b**, zone A. REE corrected to 8 wt% MgO, and normalized to the average glass composition from the two ridge segments immediately west of the MFZ. Grey dotted lines in **a** represent theoretical curves for non-modal near fractional melting at 1.2% melting per kbar (ref. 2) normalized to the theoretical liquid produced at ~5% melting. The red line in **b** represents 5% melting with half the melting in the garnet stability field normalized to a liquid produced by 5% melting in the spinel stability field only. The green line (‘fertilized residual mantle’) corresponds to the relative composition (compared to the original source) of a mantle having undergone 10% melting, then fertilized by 5 relative per cent of a liquid produced by ~5% melting. This curve is only directly comparable to that of the normalized glass data if it is assumed that the degree of melting of the refertilized source is the same as that of the composition used to normalize the glass data. Even so, other degrees of melting of the refertilized source produce qualitatively similar profiles. Any secondary clinopyroxene crystallized during fertilization will be low enough in abundance to be completely eliminated during remelting. Further details are given in Supplementary Information.

In order to constrain the origin of zone A basalts, we will concentrate on the rare-earth elements (REE), as relative fractionations between these elements are known to provide a sensitive probe of the partial melting process. For example, the different compatibility of different REE in clinopyroxene leads to a pronounced relative depletion in the light REE (LREE) in the liquids as melting proceeds, as illustrated in Fig. 4a for basalts from zone B. In this representation all liquid compositions have been corrected to 8 wt% MgO, then normalized to the average composition of glasses from the two ridge segments in zone B immediately to the west of the MFZ (Supplementary Table 4). First, this normalization eliminates the need to know the source composition as long as a single source is assumed. Second, this average was chosen because these segments reflect the lowest degree of partial melting from zone B (estimated to be ~6.7% of a typical normal-MORB source based on its Na<sub>8.0</sub> contents<sup>2</sup>), thus all other glasses from this zone should represent higher degrees of melting. As expected, normalized REE patterns are fan-shaped (Fig. 4a), with the most depleted compositions occurring at the western extremity of the zone, at the shallowest depths and having the lowest Na<sub>8.0</sub>, Ti<sub>8.0</sub> and (Sm/Yb)<sub>n</sub>, and highest Fe<sub>8.0</sub> and CaO/Al<sub>2</sub>O<sub>3</sub>; all of these melt proxies are consistent with a progressive increase in degree of partial melting of a fertile lherzolite from east to west in zone B.

In marked contrast, when zone A basalts are corrected to 8 wt% MgO and normalized to the same average composition, a completely different REE pattern is apparent (Fig. 4b). The normalized spectra are U-shaped, with a steep decrease in relative concentration from La to Sm then a slight positive slope from Sm to Yb. Another significant feature is that the spectra are all more or less parallel with little relative fractionation between REE as a function of absolute concentration, contrasting with zone B basalts (Fig. 4b). We note that the concentrations of the HREE are systematically depleted relative to the reference composition. From this representation, it is clear that basalts from zones A and B have not been produced by different degrees of partial melting of the same source. Melting in the garnet stability field, as has been previously suggested for this region<sup>16,17</sup>, can also be excluded, as significant involvement of garnet would fractionate the HREE from the LREE, in contrast to our observations (Fig. 4b).

The parallel nature of the REE patterns in zone A (Fig. 4b) is inconsistent with the presence of clinopyroxene in the residue over the sampled melting interval. We note that this does not exclude the possibility that a small proportion of clinopyroxene was present at melt fractions lower than those sampled by the basalts. The inference that clinopyroxene was absent from the residue also provides an explanation for the observed correlation between CaO/Al<sub>2</sub>O<sub>3</sub> and Ti<sub>8.0</sub> (Fig. 2b), because once clinopyroxene is exhausted during melting, experimental studies<sup>18,19</sup> have shown that CaO/Al<sub>2</sub>O<sub>3</sub> of the liquid changes from increasing to decreasing values. Independent support for this idea is provided by peridotites dredged from zone A, which are notably poor in clinopyroxene<sup>20</sup>. The simplest way to produce a mantle depleted in clinopyroxene is for it to have undergone a prior partial melting event, a hypothesis consistent with the strong depletion in HREE (Fig. 4b). However, this simple model cannot explain the high concentrations of the more incompatible elements such as the LREE or Na, which should have been almost entirely removed from the source during the first melting event, and some re-fertilization of this depleted source is required. Although the exact nature of this contaminating liquid and its origin cannot be determined with certainty, the REE pattern of zone A liquids can be satisfactorily accounted for by mixture of a harzburgite and a partial melt of a fertile peridotite (Fig. 4b).

This interpretation begs the question of whether the SWIR has the attributes of a 'cold section' because mantle potential temperature is low, or because the mantle source is less fertile and thus less able to produce melt. Furthermore, it raises questions concerning the

geodynamic origins of zone A basalts. Previous melting of the asthenosphere beneath the central Indian ridge may play a part (Fig. 1), but this cannot explain the fact that the mantle has been affected over distances of ~1,000 km. Alternatively these basalts may reflect melting of ultra-depleted asthenosphere that has been entrained into the ridge system and is floating above 'normal' asthenosphere. Such depleted sources may be a general feature of the Indian Ocean mantle, resulting in a different spectrum of source composition than present beneath the Atlantic and Pacific oceans. □

Received 9 September 2002; accepted 9 January 2003; doi:10.1038/nature01424.

1. Klein, E. M. & Langmuir, C. H. Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
2. Langmuir, C. H., Klein, E. M. & Plank, T. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps-Morgan, J., Blackmann, D. K. & Sinton, J. M.) 183–280 Geophysical Monograph series, Vol. 71 (American Geophysical Union, Washington DC, 1992).
3. Humler, E., Thiriot, J. L. & Montagner, J. P. Global correlations of mid-ocean-ridge basalt chemistry with seismic tomographic images. *Nature* **364**, 225–228 (1993).
4. Cannat, M., Rommevaux-Jestin, C., Sauter, D., Deplus, C. & Mendel, V. Formation of the axial relief at the very slow spreading Southwest Indian Ridge (49°–69°E). *J. Geophys. Res.* **104**, 22825–22843 (1999).
5. Debayle, E. & LeVèque, J. J. Upper mantle heterogeneities in the Indian Ocean from waveform inversion. *Geophys. Res. Lett.* **24**, 245–248 (1997).
6. Muller, M. R., Minshull, T. A. & White, R. S. Segmentation and melt supply at the Southwest Indian Ridge. *Geology* **27**, 867–870 (1999).
7. Shu, D. & Gordon, R. G. Evidence for motion between Nubia and Somalia along the Southwest Indian Ridge. *Nature* **398**, 64–66 (1999).
8. Schilling, J. G. *et al.* Thermal structure of the mantle beneath the Equatorial Mid-Atlantic Ridge: Inferences from the spatial variation of dredged basalt glass compositions. *J. Geophys. Res.* **100**, 10057–10076 (1995).
9. Bryan, W. B., Humphris, S. E., Thompson, G. & Casey, J. F. Comparative volcanology of small axial eruptive centers in the MARK area. *J. Geophys. Res.* **99**, 2973–2984 (1994).
10. Natland, J. H. in *Oceanic Basalts* (ed. Floyd, P. A.) 288–310 (Blackie and Son, Glasgow, 1991).
11. Elthon, D., Ross, D. K. & Meen, J. K. Compositional variations of basaltic glasses from the Mid-Cayman Rise spreading center. *J. Geophys. Res.* **100**, 12497–12512 (1995).
12. Elthon, D., Ross, D. K. & Meen, J. K. Correction to "Compositional variations of basaltic glasses from the Mid-Cayman Rise spreading center". *J. Geophys. Res.* **101**, 17577–17579 (1996).
13. Klein, E. M., Langmuir, C. H. & Staudigel, H. Geochemistry of basalts from the Southeast Indian Ridge 115°–138°E. *J. Geophys. Res.* **96**, 2089–2107 (1991).
14. Pyle, D. G., Christie, D. M., Mahoney, J. J. & Duncan, R. A. Geochemistry and geochronology of ancient southeast Indian and southwest Pacific seafloor. *J. Geophys. Res.* **100**, 22261–22282 (1995).
15. Pyle, D. G. *Geochemistry of Mid-ocean Ridge Basalt Within and Surrounding the Australian-Antarctic Discordance*. Thesis, Oregon State Univ. (1994).
16. White, R. S., Minshull, T. A., Bickle, M. J. & Robinson, C. J. Melt generation at very slow-spreading oceanic ridges: Constraints from geochemical and geophysical data. *J. Petrol.* **42**, 1171–1196 (2001).
17. Robinson, C. J., Bickle, M. J., Minshull, T. A., White, R. S. & Nichols, A. R. L. Low degree melting under the Southwest Indian Ridge: the roles of mantle temperature, conductive cooling and wet melting. *Earth Planet. Sci. Lett.* **188**, 383–398 (2001).
18. Pickering-Witter, J. & Johnston, A. D. The effects of variable bulk composition on the melting systematics of fertile peridotitic assemblages. *Contrib. Mineral. Petrol.* **140**, 190–211 (2000).
19. Falloon, T. J. & Danyushevsky, L. V. Melting of refractory mantle at 1.5, 2 and 2.5 GPa under anhydrous and H<sub>2</sub>O-undersaturated conditions: Implications for the petrogenesis of high-Ca boninites and influence of subduction components on mantle melting. *J. Petrol.* **41**, 257–283 (2000).
20. Seyler, M., Cannat, M. & Mével, C. Evidence for major-element heterogeneity in the mantle source of abyssal peridotites from the Southwest Indian Ridge (52° to 68°E). *Geochem. Geophys. Geosyst.* (in the press).
21. Smith, W. H. F. & Sandwell, D. T. Global seafloor topography from satellite altimetry and ship depth soundings. *Science* **277**, 1957–1962 (1997).
22. Patriat, P., Sauter, D., Munschy, M. & Parson, L. A survey of the Southwest Indian Ridge axis between Atlantis P fracture zone and the Indian Ocean triple junction: regional setting and large-scale segmentation. *Mar. Geophys. Res.* **19**, 457–480 (1997).
23. Plank, T. & Langmuir, C. H. Effects of the melting regime on the composition of the oceanic crust. *J. Geophys. Res.* **97**, 19749–19770 (1992).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com/nature>).

**Acknowledgements** We thank the Service d'Analyse des Roches et des Minéraux (SARM) and J. Carignan for their help in the laboratory. We also thank D. Sauter for helpful discussions. Funding was provided by CNRS-INSU, and the IFRTF provided access to the Marion Dufresne II for sampling.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to C.M. (e-mail: cm@dlc.ku.dk).



# Single origin of Malagasy Carnivora from an African ancestor

Anne D. Yoder<sup>\*†</sup>, Melissa M. Burns<sup>†</sup>, Sarah Zehr<sup>‡</sup>, Thomas Delefosse<sup>\*</sup>, Geraldine Veron<sup>§</sup>, Steven M. Goodman<sup>†||</sup> & John J. Flynn<sup>‡</sup>

<sup>\*</sup> Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, Connecticut 06511, USA

<sup>†</sup> Department of Zoology, and <sup>‡</sup> Department of Geology, Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, Illinois 60605, USA

<sup>§</sup> Laboratoire de Zoologie (Mammifères et Oiseaux), Muséum National d'Histoire Naturelle, Paris, France

<sup>||</sup> WWF, B.P. 738, Antananarivo (101), Madagascar

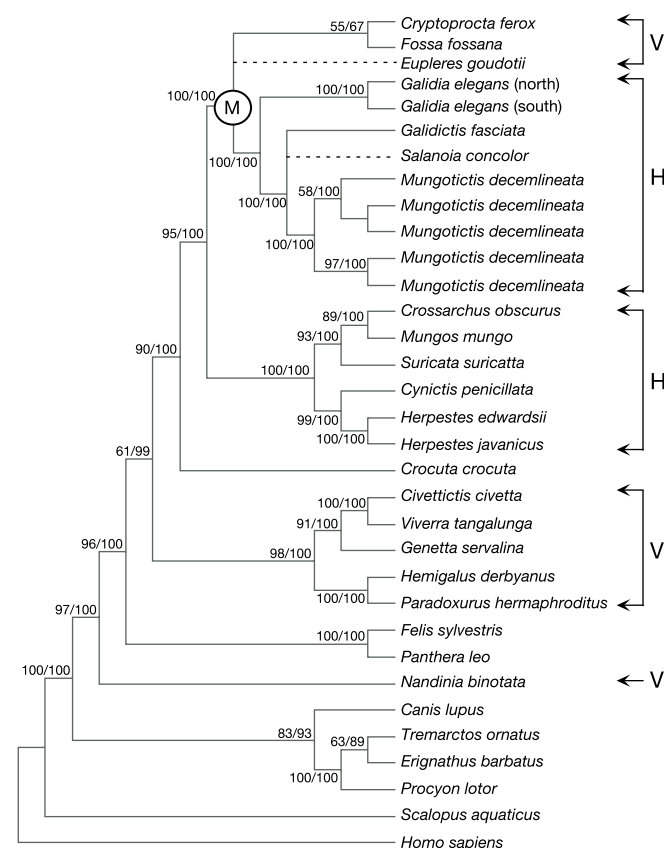
The Carnivora are one of only four orders of terrestrial mammals living in Madagascar today. All four (carnivorans, primates, rodents and lipotyphlan insectivores) are placental mammals with limited means for dispersal, yet they occur on a large island that has been surrounded by a formidable oceanic barrier for at least 88 million years<sup>1,2</sup>, predating the age of origin for any of these groups<sup>3,4</sup>. Even so, as many as four colonizations of Madagascar have been proposed for the Carnivora alone<sup>5</sup>. The mystery of the island's mammalian origins is confounded by its poor Tertiary fossil record, which leaves us with no direct means for estimating dates of initial diversification. Here we use a multi-gene phylogenetic analysis to show that Malagasy carnivorans are monophyletic and thus the product of a single colonization of Madagascar by an African ancestor. Furthermore, a bayesian analysis<sup>6</sup> of divergence ages for Malagasy carnivorans and lemuriforms indicates that their respective colonizations were temporally separated by tens of millions of years. We therefore conclude that a single event, such as vicariance or common dispersal, cannot explain the presence of both groups in Madagascar.

Madagascar is an island of remarkable faunal and floral diversity where species endemism reaches 100% in some animal groups. It is also unusual for its endemism at high taxonomic levels and skewed representation of many biotic groups compared with continental communities. Krause *et al.*<sup>3</sup> have referred to these patterns as Madagascar's unique signature of 'imbalance and endemism'. This signature undoubtedly derives from the island's long history of complete isolation from other landmasses. Madagascar, and other fragments of the Gondwanan supercontinent, split from mainland Africa nearly 165 million years (Myr) ago and began its south-eastward trajectory to reach its current position at approximately 121 Myr ago (ref. 1). For this entire period, and the subsequent 30 Myr or so, it remained attached to India, forming the Indo-Madagascar continent. These two landmasses finally separated approximately 88 Myr ago (ref. 2) when India started its northward movement to subsequently collide with the Asian continent in the Early Eocene<sup>7</sup>. Thus, although Madagascar had a long association with India in particular, and remains within 400 km of Africa, it has otherwise been completely isolated and surrounded by a deep and wide oceanic barrier for at least the past 88 Myr. Although it has been proposed that a nearly continuous landbridge linked Madagascar to Africa for much of the Cenozoic<sup>8</sup>, the geologic data on which this hypothesis is based are extremely frail<sup>9</sup>.

The various biogeographical mechanisms that could explain the presence of mammals and other terrestrial vertebrates in Madagascar are: Gondwanan vicariance<sup>10</sup>; land-mediated dispersal (for example, landbridge connections)<sup>8</sup>; and 'sweepstakes' over-water dispersal (that is, 'rafting')<sup>11,12</sup> or island-hopping<sup>3</sup>. Each model is compatible with contrasting phylogenetic and temporal patterns. The Gondwanan vicariance mechanism implies ancient divergence dates, with all clades of that antiquity having equally probable

chances of occurrence in Madagascar. The landbridge model<sup>8</sup> implies divergence dates constrained within a period from the Middle Eocene epoch (~45 Myr ago) until the Early Miocene epoch (~26 Myr ago), also with equally probable opportunities for clade representation. The sweepstakes dispersal model, however, implies that divergence dates should be more randomly distributed within the time frame of placental evolution, and that some groups would be more suited to withstand over-water dispersal than others. Ascertaining which of these mechanisms appears to have been the most likely requires determining both the phylogeny and the age of Malagasy lineages<sup>8</sup>.

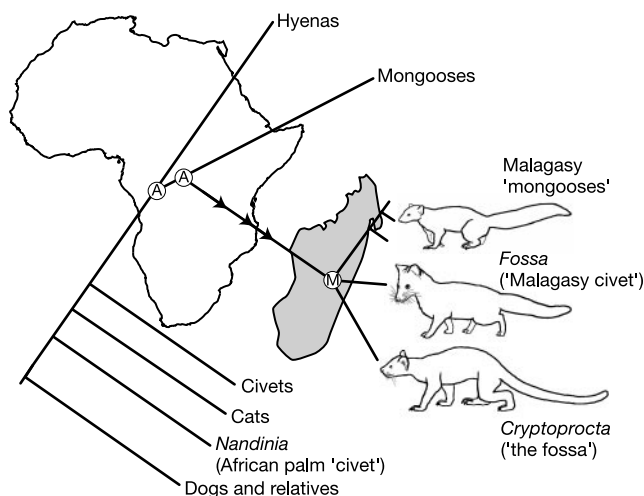
Using maximum likelihood, minimum evolution, maximum parsimony and bayesian analysis, we phylogenetically analysed the Malagasy Carnivora and related feliform and caniform outgroups by sampling four genes from three unlinked genetic loci. The non-carnivoran outgroups, human (*Homo sapiens*) and mole (*Scalopus aquaticus*), were included to ensure proper rooting of the phylogeny. Several notable results emerge from the phylogenetic analysis, regardless of the optimality criteria used. The Malagasy carnivorans are shown to constitute a clade with robust statistical support (Fig. 1), which is most relevant for testing the validity of the alternative biogeographical models. This result indicates that the



**Figure 1** Maximum likelihood phylogeny of Malagasy Carnivora and selected carnivoran outgroups. The tree is rooted with mole (*Scalopus aquaticus*) and human (*Homo sapiens*). The tree topology has a likelihood score of 38580.775. The dashed lines indicate taxa sampled from museum specimens for which data could only be obtained from the cytochrome b gene because of poor DNA quality. The tree otherwise represents an analysis wherein the four data sets were combined in a single analysis. Numbers on nodes represent statistical support from a 'fast' bootstrap analysis under the likelihood criterion (first number) and posterior probabilities from the bayesian analysis (second number). Circled M, Malagasy carnivoran node; V, taxa previously classified within the family Viverridae; H, within the family Herpestidae. (See also Table in Supplementary Information.)

Malagasy Carnivora are the product of a single colonization of Madagascar, a conclusion that is consistent with phylogenetic analyses of the other three groups of endemic modern terrestrial mammals in Madagascar<sup>13–15</sup>. Within the Malagasy Carnivora clade, *Cryptoprocta* and *Fossa* appear to be basal (although their relative relationships within the Malagasy clade are poorly resolved, as indicated by weak statistical support and/or differing resolutions depending on optimality criteria) and the mongoose-like lineages (*Galidia*, *Galidictis* and *Mungotictis*) are shown to comprise a well-supported internal clade. Cytochrome *b* data from museum skins of *Salanoia* and *Eupleres* indicate that they too lie within the Malagasy carnivoran clade, with *Eupleres* falling in an unresolved basal position and *Salanoia* belonging to the Malagasy ‘mongoose’ sub-clade. A clade comprising African and Asian Herpestidae is found by all optimality criteria to be the nearest outgroup of the Malagasy clade, with the African Hyaenidae as sister to this combined clade. Together, these results clearly indicate that the Malagasy carnivorans, including the large and very puma-like *Cryptoprocta*<sup>16</sup>, are most closely related to the Herpestidae (in the strict sense), which are optimized to be of African origin (Fig. 2). In the higher-level systematics of the Carnivora, the Malagasy clade lies deeply nested within a feliform clade with the ‘viverrid’ *Nandinia* as the basal feliform lineage. The Malagasy clade does not, however, belong to a monophyletic Viverridae, nor do subsets of Malagasy taxa nest within the Herpestidae or Felidae, as has been previously postulated<sup>5</sup>. Instead, our results further confirm that the traditional family Viverridae is paraphyletic<sup>17</sup> and, contrary to long-held beliefs, does not contain any Malagasy species.

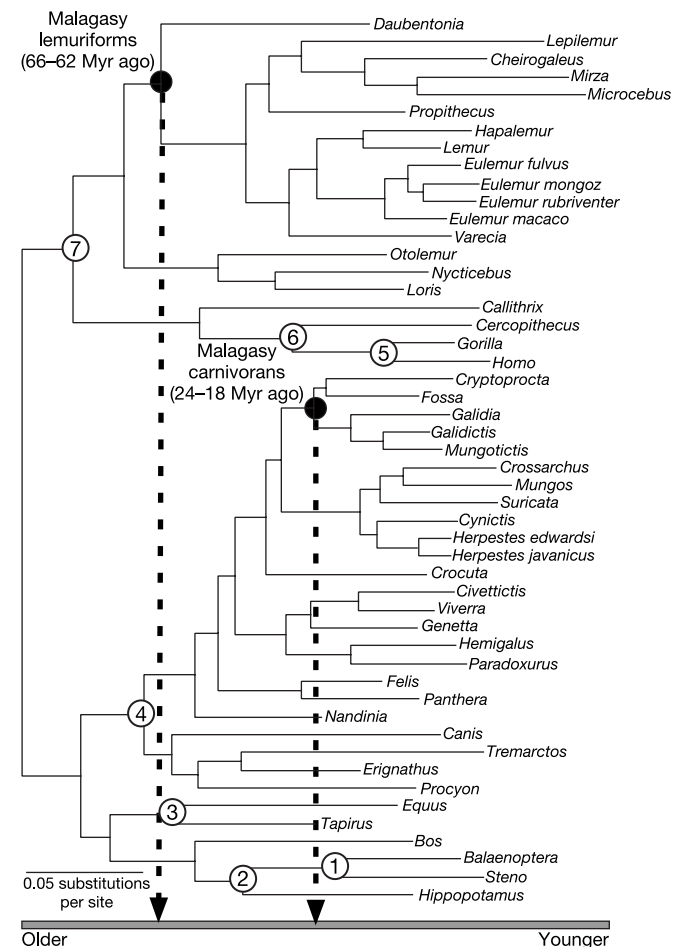
Given that the hypothesis of a single origin from an African ancestor for Malagasy carnivorans is precisely congruent with similar phylogeographical observations of Malagasy Primates (Lemuriformes)<sup>15,18</sup>, it becomes informative to investigate the relative divergence times of the two groups. If a single biogeographical event, such as vicariance or common dispersal, were to explain the presence of both groups in Madagascar, we would expect to see similar (if not identical) age estimates for both. We used bayesian methods<sup>6</sup> to estimate divergence ages for the Malagasy carnivorans and lemuriforms. These methods do not assume a molecular clock and also allow for several fossil-based calibrations



**Figure 2** Proposed biogeographical model of carnivoran dispersal from Africa. The figure illustrates single colonization of Madagascar and subsequent radiation into extant lineages. A, African radiation; M, Malagasy radiation. Malagasy radiation is illustrated as a trifurcation because of the poor resolution (that is, low statistical support and/or conflicting placements) of *Cryptoprocta* and *Fossa* with respect to the Malagasy ‘mongoose’ clade. The placement of *Salanoia* and *Eupleres* is not illustrated or implied because of limited data for these taxa. (Figure prepared by M. Hill Donnelly, Field Museum of Natural History.)

to be used simultaneously in a given analysis. Dates for the Malagasy Carnivora were estimated individually for the four genes sampled by this study and for a data set in which the four genes were combined. For the *interphotoreceptor retinoid-binding protein* (IRBP) and *cytochrome b* data sets, and for their combination, an extensive sample of lemuriform taxa was available, allowing a simultaneous analysis of the two groups. This latter comparison is important because it uses homologous data and identical methods of analysis for two independent clades of Madagascar’s colonizing mammals. Consequently, even if there is error in the determination of the absolute age of the two clades, the determination of relative ages should be similarly biased and therefore directly comparable.

Visual inspection of the phylogeny used for the comparative age analysis (Fig. 3), wherein branches are drawn to be proportional to estimated branch lengths, shows an obvious differential in clade depth between the two groups. This differential is clarified by the bayesian analysis, which indicates that the two Malagasy clades are



**Figure 3** Comparative age analysis of Lemuriformes and Malagasy Carnivora. Branch lengths from combined IRBP and cytochrome *b* tree are drawn to be proportional with estimated rates of change per site as calculated by *estbranches*. The tree was rooted with *Didelphis virginiana* and *Mus musculus* (not shown). The time bar at the bottom of the figure is not meant to imply a clock-like tree. Numbered nodes indicate calibration points in the bayesian analysis. These are: (1) toothed whale/baleen whale divergence 33–40 Myr ago; (2) whale/hippo divergence 51–60 Myr ago; (3) horse/tapir (or rhinoceros for ND2) divergence 50–58 Myr ago; (4) caniform/feliform divergence 45–65 Myr ago; (5) human/gorilla divergence 8–12 Myr ago; (6) monkey/ape divergence 32–38 Myr ago; (7) basal primate radiation 63–90 Myr ago. Owing to poor outgroup sequence availability for transthyretin, only calibrations 4, 5 and 6 were used in the analysis of that gene, and calibration 4 in the four-gene combined analysis. See Supplementary Information for relevant literature citations.

Table 1 Bayesian estimates of geologic age by data set

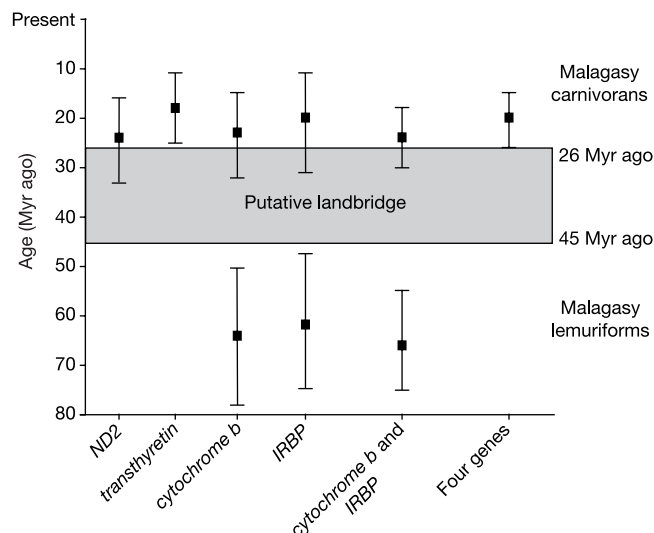
| Data set                                | Malagasy Carnivora | Malagasy Primates  |
|-----------------------------------------|--------------------|--------------------|
| <i>Cytochrome b</i> (1,140 bp)          | 23 Myr ago (15–32) | 64 Myr ago (50–78) |
| <i>IRBP</i> (exon 1, 945 bp)            | 20 Myr ago (11–31) | 62 Myr ago (47–75) |
| <i>ND2</i> (1,044 bp)                   | 24 Myr ago (16–33) | NA                 |
| <i>Transthyretin</i> (intron 1, 897 bp) | 18 Myr ago (11–25) | NA                 |
| <i>Cytochrome b</i> and <i>IRBP</i>     | 24 Myr ago (18–31) | 66 Myr ago (55–75) |
| Four genes combined                     | 20 Myr ago (15–26) | NA                 |

See Fig. 3 legend for description of fossil calibrations used for the estimation of geologic ages. Numbers in parentheses represent 95% credibility intervals. NA, not available.

of considerably different ages (Table 1). The range of posterior mean estimates for the Malagasy carnivorans lies between 24 and 18 Myr ago, whereas that for the lemuriforms lies between 66 and 62 Myr ago, suggesting that their relative colonizations were separated by 38 Myr. Even when one examines the extreme boundaries of the 95% credibility intervals (95% CI) for the two groups, the absolute minimum time separating their relative divergence ages is 14 Myr (or, likewise, as many as 67 Myr). Most importantly, in any of the possible comparisons, the disparity in age estimates indicates that the same biogeographical event cannot explain the joint appearance of both groups in Madagascar.

What then are the biogeographical implications of this temporal analysis? If the presence of either or both groups were to be explained by Indo-Madagascar vicariance, we would expect to see substantially more ancient divergence ages (namely, >88 Myr), with each group having its closest relatives in Indo-Asia. We would also expect to see representation of closely related groups (for example, other feliforms, caniform carnivorans, lorisiform primates) in Madagascar. This is not what we observe. Furthermore, the fact that both Malagasy clades appear to have origins in the Tertiary period is compatible with the known fossil record of Madagascar. The only mammals known from the Late Cretaceous period of Madagascar are multituberculates, gondwanatheres and marsupials<sup>19,20</sup>, all of which are now extinct on the island. Conversely, colonization of Madagascar via the hypothesized landbridge would imply posterior mean-age estimates that were approximately equivalent in the two groups, with ages for both falling within the proposed temporal window. Moreover, we would expect to see greater representation of both closely and distantly related groups of mammals which also would have found dispersal across a continuous landbridge to be expedient. Again, these are not the observed patterns. Instead, the data are most compatible with a model of over-water dispersal from African ancestry. All of the posterior mean-age estimates for both clades are entirely outside the 45–26 Myr ago temporal window hypothesized for the putative landbridge (Fig. 4), which is actually rather remarkable given that this window spans more than one-third of the Cenozoic era. Although the extreme lower bounds of the 95% CIs for the carnivorans overlap slightly with the temporal window, the consistent estimation of posterior means between 18–24 Myr ago from four different genes suggests that they are a good approximation of actual divergence age, and all post-date the entire span of the putative landbridge. Finally, it is notable that the 95% CIs are reduced in both of the combined data analyses, and in the four-gene combination for the carnivorans, do not overlap with the landbridge temporal window.

As of late 2001, the total known extant native terrestrial mammal fauna of Madagascar comprises 101 species (ref. 21) belonging to four independent clades (rodents<sup>13</sup>, lipotyphlans<sup>14</sup>, lemurs<sup>15</sup> and carnivorans—herein). Thus, the entire extant land mammal fauna of the island can be explained by only four colonization events, and at least for the two clades compared here, their immigrations appear to have occurred at random as predicted by Simpson's 'sweepstakes' model of over-water dispersal<sup>12</sup>. It is significant to note that for at least three of the four founding clades of terrestrial mammals, some



**Figure 4** Comparison of Bayesian age estimates for Malagasy carnivorans (above) and lemuriforms (below) relative to proposed landbridge temporal window<sup>8</sup>. Filled squares, posterior mean estimates of clade divergence ages in Madagascar; thin lines, 95% credibility intervals for age estimates.

species have ecophysiological specializations, such as the ability to hibernate or go into torpor<sup>22,23</sup>, which would have predisposed them for enduring long periods of drought and food shortage<sup>11,24,25</sup>. If these specializations are primitive within each group, as has been concluded for the tenrecs<sup>14</sup>, the over-water dispersal model gains yet more credence. Given the morphological and ecological diversity of the island's existing native mammals, this further underscores that the successful crossing of a large water barrier, with subsequent colonization and diversification, has been a very rare event indeed among non-volant mammals. This pattern makes for a comparison with that observed for chameleons, wherein over-water dispersal has also been posited as the likely mechanism explaining their biogeographical distribution<sup>26</sup>. In this group, however, 'sweepstakes' dispersal has been inferred to have been a relatively common event, in contrast to the extreme rarity here reported for terrestrial mammals. □

## Methods

Two mitochondrial genes, *cytochrome b* and *ND2*, were sequenced in their entirety, yielding alignments of 1,140 base pairs (bp) and 1,044 bp, respectively. Two nuclear genes were partly sequenced: exon 1 of the *interphotoreceptor retinoid-binding protein* (*IRBP*), and intron 1 of the *transthyretin* gene, yielding alignments of 945 bp and 897 bp, respectively. In humans, *transthyretin* is located on chromosome 18 and *IRBP* on chromosome 10; we thus assume that they are also independently segregating in Carnivora. All sequences were aligned by eye, owing to a lack of indels, except for the *transthyretin* sequences. For this gene, sequences were aligned with Clustal W (1.4) and adjusted manually. Before phylogenetic analysis, all large taxon-specific indels were omitted. Phylogenetic analysis was done on a combined data set of the four genes with PAUP\* version 4.0b6 (Altivec)<sup>27</sup> by using the maximum parsimony, minimum evolution and maximum likelihood optimality criteria. Settings were as follows: maximum parsimony with starting trees obtained by stepwise addition, minimum evolution with an HKY85 correction and starting trees obtained via neighbour-joining, and maximum likelihood with a GTR +  $\Gamma$  + I model; all other settings were set by default. To assess statistical support for hypothesized clades, bootstrap analysis was done with 100 bootstrap replicates sampling 10 replicates of the random stepwise addition option; 100 replicates of the 'fast' bootstrap algorithm was used for the maximum likelihood analysis. Sequences were also analysed with a Markov chain Monte Carlo approach using the program MrBayes version 2.01<sup>28</sup> with clade support assessed by posterior probability. Four Markov chains, one cold and three heated, were allowed to run for 1,000,000 generations using random starting trees. The model used was equivalent to a GTR +  $\Gamma$  likelihood model. MacClade 4.0 PPC<sup>29</sup> was used to optimize geographical origins for internal nodes within the feliform carnivorans.

Appropriate likelihood models for the Bayesian analysis of divergence ages were determined using the program MODELTEST version 3.06 PPC<sup>30</sup> with the mean rate for each category of the discrete approximation of the gamma distribution estimated with PAUP\*. The program *estbranches* was used to estimate branch lengths for the *IRBP*.



*cytochrome b*, *ND2* and *transferrin* genes individually and in various combinations (see text). We used the program *divtime5bmac* to estimate divergence times, using the output from *estbranches* (both programs are available from J. Thorne). Markov chain Monte Carlo analyses were run for 1 million generations after a burnin of 100,000 generations. Chains were sampled every 100 generations, yielding 10,000 samples of the Markov chain. Seven independent calibrations were used specifying upper and lower bounds (see Fig. 3). For each analysis, one approximation of the prior distribution and at least two separate approximations of the posterior distribution were obtained. Comparison of the prior and posterior approximations reveals the impact of the sequence data on the divergence time estimates. The separate approximations of the posterior distributions were done by starting the Markov chain Monte Carlo technique from different initial states. The similarity of these approximations indicates that the Markov chains have successfully converged to the posterior distribution.

Sequences are deposited in GenBank under accession numbers AY170012–AY170116, AY187007, and AY187008. Sequence alignments are available in Supplementary Information.

Received 29 May; accepted 5 November 2002; doi:10.1038/nature01303.

- Coffin, M. F. & Rabinowitz, P. D. in *Geology and Geophysics of Continental Margins* (eds Watkins, J. S., Zhiqiang, F. & McMillen, K. J.) 207–246 (American Association of Petroleum Geologists, Tulsa, Oklahoma, 1992).
- Storey, M. *et al.* Timing of hot-spot related volcanism and the breakup of Madagascar and India. *Science* **267**, 852–855 (1995).
- Krause, D. W., Hartman, J. H. & Wells, N. A. in *Natural Change and Human Impact in Madagascar* (eds Goodman, S. M. & Patterson, B. D.) 3–43 (Smithsonian Institution Press, Washington DC, 1997).
- Novacek, M. J. Mammalian phylogeny: Genes and supertrees. *Curr. Biol.* **11**, R573–R575 (2001).
- Veron, G. La position systématique de *Cryptoprocta ferox* (Carnivora). Analyse cladistique des caractères morphologiques de carnivores Aeluroidea actuels et fossiles. *Mammalia* **59**, 551–582 (1995).
- Thorne, J. L., Kishino, H. & Painter, I. S. Estimating the rate of evolution of the rate of evolution. *Mol. Biol. Evol.* **15**, 1647–1657 (1998).
- Rowley, D. B. Age of initiation of collision between India and Asia: a review of stratigraphic data. *Earth Planet. Sci. Lett.* **145**, 1–13 (1996).
- McCall, R. A. Implications of recent geological investigations of the Mozambique Channel for the mammalian colonization of Madagascar. *Proc. R. Soc. Lond. B* **264**, 663–665 (1997).
- Rogers, R. R., Hartman, J. H. & Krause, D. W. Stratigraphic analysis of Upper Cretaceous rocks in the Mahajanga Basin, northwestern Madagascar: implications for ancient and modern faunas. *J. Geol.* **108**, 275–301 (2000).
- Arnason, U., Gullberg, A., Burguete, A. S. & Janke, A. Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. *Heredity* **133**, 217–228 (2000).
- Martin, R. D. Adaptive radiation and behaviour of Malagasy lemurs. *Phil. Trans. R. Soc. Lond. B* **264**, 295–352 (1972).
- Simpson, G. G. Probabilities of dispersal in geologic time. *Bull. Am. Mus. Nat. Hist.* **99**, 163–176 (1952).
- Jansa, S. A., Goodman, S. M. & Tucker, P. K. Molecular phylogeny and biogeography of the native rodents of Madagascar (Muridae: Nesomyinae): a test of the single-origin hypothesis. *Cladistics* **15**, 253–270 (1999).
- Olson, L. E. *Systematics, Evolution, and Biogeography of Madagascar's Tenrecs (Mammalia: Tenrecidae)*. 349, PhD Thesis, Univ. Chicago (1999).
- Yoder, A. D., Cartmill, M., Ruvelo, M., Smith, K. & Vilgalys, R. Ancient single origin of Malagasy primates. *Proc. Natl Acad. Sci. USA* **93**, 5122–5126 (1996).
- Veron, G. & Catzeflis, F. M. Phylogenetic relationships of the endemic Malagasy carnivore *Cryptoprocta ferox* (Aeluroidea): DNA/DNA hybridization experiments. *J. Mamm. Evol.* **1**, 169–185 (1993).
- Flynn, J. J. & Nedbal, M. A. Phylogeny of the Carnivora (Mammalia): congruence vs. incompatibility among multiple data sets. *Mol. Phylogenet. Evol.* **9**, 414–426 (1998).
- Martin, R. D. Origins, diversity, and relationships of lemurs. *Int. J. Primatol.* **21**, 1021–1049 (2000).
- Krause, D. W. *et al.* The Late Cretaceous vertebrate fauna of Madagascar: Implications for Gondwanan paleobiogeography. *Geol. Soc. Am. Today* **9**, 1–7 (1999).
- Krause, D. W. Fossil molar from a Madagascan marsupial. *Nature* **412**, 497–498 (2001).
- Goodman, S. M., Ganzhorn, J. & Rakotondravony, D. in *Natural History of Madagascar* (eds Goodman, S. M. & Benstead, J.) (Univ. Chicago Press, Chicago, in the press).
- Albignac, R. *Faune de Madagascar*. 36 Mammifères carnivores, 206 (ORSTOM/CNRS, Paris, 1973).
- Schmid, J. & Stephenson, P. J. in *Natural History of Madagascar* (eds Goodman, S. M. & Benstead, J.) (Univ. Chicago Press, Chicago, in the press).
- Kappeler, P. M. Lemur origins: rafting by groups of hibernators? *Folia Primatol. (Basel)* **71**, 422–425 (2000).
- Yoder, A. D. in *Biogéographie de Madagascar* (ed. Lourenco, W. R.) 245–258 (Editions de l'ORSTOM, Paris, 1996).
- Raxworthy, C. J., Forstner, M. R. & Nussbaum, R. A. Chameleon radiation by oceanic dispersal. *Nature* **415**, 784–787 (2002).
- Swofford, D. L. PAUP\*: phylogenetic analysis using parsimony (\*and other methods) (Sinauer Associates, Sunderland, Massachusetts, 1998).
- Huelsenbeck, J. P. & Ronquist, F. MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
- Maddison, W. P. & Maddison, D. R. MacClade: analysis of phylogeny and character evolution (Sinauer Associates, Sunderland, Massachusetts, 1992).
- Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com/nature>).

**Acknowledgements** We thank the authorities in Madagascar and the WWF for support of fieldwork, and L. Razafimanantsoa for access to tissues. We thank Z. Yang and M. Springer for discussion of methods, D. Krause for comments, and J. Thorne for guidance with his programs. This work was supported by the National Science Foundation, Washington, and the Field Museum's Pritzker Laboratory for Molecular Systematics. This is a Duke University Primate Center publication.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to A.D.Y. (e-mail: anne.yoder@yale.edu).

## The effect of aggressiveness on the population dynamics of a territorial bird

F. Mougeot<sup>\*†</sup>, S. M. Redpath<sup>\*†</sup>, F. Leckie<sup>\*</sup> & P. J. Hudson<sup>‡§</sup>

<sup>\*</sup> CEH Banchory, Hill of Brathens, Glassel, Banchory, Aberdeenshire AB31 4BW, UK

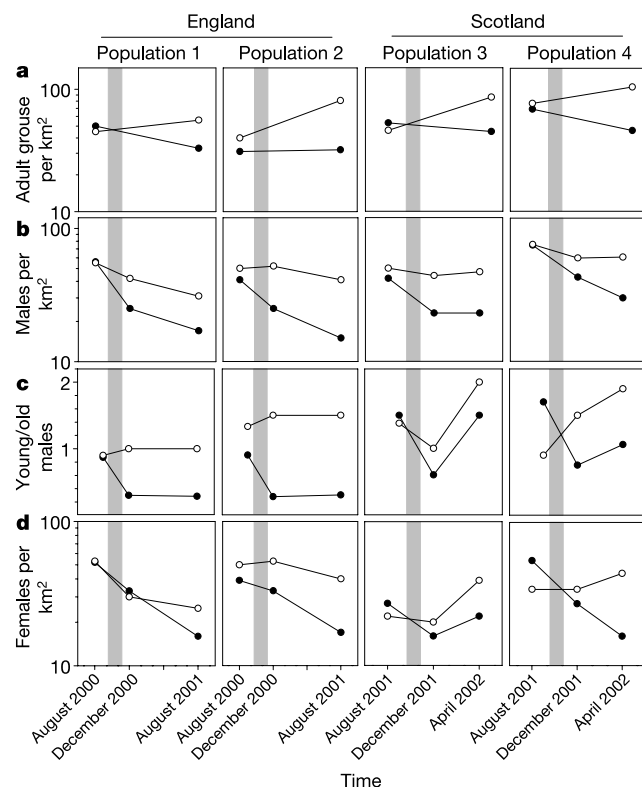
<sup>‡</sup> Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, UK

<sup>†</sup> These authors contributed equally to this work

A central issue in ecology lies in identifying the importance of resources, natural enemies and behaviour in the regulation of animal populations. Much of the debate on this subject has focused on animals that show cyclic fluctuations in abundance<sup>1–7</sup>. However, there is still disagreement about the role of extrinsic (food, parasites or predators) and intrinsic (behaviour) factors in causing cycles<sup>2,8–10</sup>. Recent studies have examined the impact of natural enemies<sup>1,3,4,7</sup>, although spatial patterns resulting from restricted dispersal or recruitment are increasingly recognized as having the potential to influence unstable population dynamics<sup>5,6,11–13</sup>. We tested the hypothesis that population cycles in a territorial bird, red grouse *Lagopus lagopus scoticus*, are caused by delayed density-dependent changes in the aggressiveness and spacing behaviour of males. Here we show that increasing aggressiveness experimentally for a short period in autumn reduced recruitment and subsequent breeding density by 50%, and changed population trajectories from increasing to declining. Intrinsic processes can therefore have fundamental effects on population dynamics.

We examined the role of intrinsic processes in the population dynamics of *Lagopus lagopus scoticus*. According to the 'territorial behaviour' hypothesis, the population cycles of this species are caused by changes in aggressiveness that influence the recruitment rate of young males in autumn<sup>14–16</sup>. Aggressiveness limits density by affecting territory size<sup>17</sup>, and shows delayed density-dependence, with a peak during the first year of a population decline<sup>15,16</sup>. The mechanism proposed for this time lag is based on changes in kinship and differential aggressive behaviour towards kin and non-kin<sup>14,15,18,19</sup>. We designed an experiment to test the effect of aggressiveness on red grouse population dynamics. Our aim was to put grouse populations at the stage in the cycle when aggressiveness is higher than expected for the prevailing density<sup>15</sup>, and to test the prediction that this would cause a population decline. We used testosterone implants to increase the aggressiveness of old (territorial) males in four populations for three months during autumn. We specifically predicted that, relative to control populations, 'aggress-

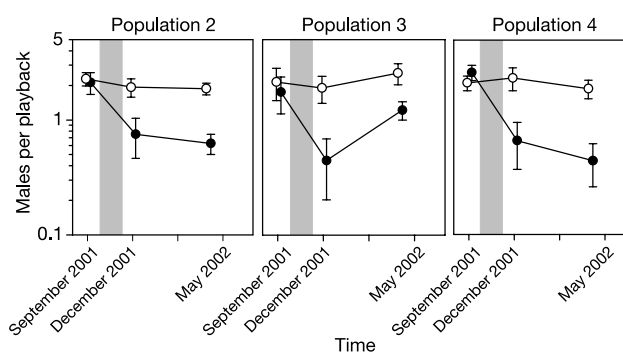
<sup>§</sup> Present address: Biology Department, Mueller Lab, Penn State University, University Park, Pennsylvania 16802, USA.



**Figure 1** Changes in grouse numbers on the control (open circles) and testosterone (filled circles) experimental areas of each population. **a**, Adult grouse per km<sup>2</sup>. **b**, Males per km<sup>2</sup>. **c**, Ratio of young to old males. **d**, Females per km<sup>2</sup>. The vertical grey bar indicates when males were caught and received implants. Note that, in part **a**, the August counts before implanting include only adult grouse (breeding birds), whereas those in parts **b** and **d** include both adult and young grouse (birds available for recruitment).

sive' populations would show, first, reduced male density by the end of autumn; second, reduced recruitment of young males; third, reduced male density in the following spring; and fourth, reduced breeding female density, in line with the reduction in male density.

Adult grouse density increased on the control areas in all four populations, whereas testosterone treatment caused population declines (populations 1, 3 and 4) or prevented the increase (population 2; Fig. 1a). The four predictions of the territorial behaviour hypothesis all held. First, the decline in male density from August to December was greater on the testosterone areas ( $45 \pm 7\%$ ) than on the control areas ( $11 \pm 14\%$ ; Fig. 1b; Table 1). This pattern was corroborated by independent playback methods for estimating male density (Fig. 2; Table 1). Second, the ratio of young to old males



**Figure 2** Changes in the mean ( $\pm$ s.e.m.) number of males contacted per playback on the control (open circles) and testosterone (filled circles) experimental areas of each study population (all  $n = 9$ ).

decreased during autumn by  $64 \pm 7\%$  on the treated areas, as compared with  $1 \pm 22\%$  on the control areas (Fig. 1c; Table 1). This indicated that aggressive, old males prevented young males from recruiting, leading to fewer young males in the autumn territorial populations. Third, the differences in male density and young-to-old ratios observed between the testosterone and control areas held until the following spring (populations 2, 3 and 4) and summer (populations 1 and 2; Fig. 1b, c; Fig. 2). Male density was then  $53 \pm 8\%$  lower on the treated than on the control areas (Fig. 1b). Thus, there was no evidence that males excluded during the treatment returned to establish territories once the testosterone implants had been exhausted, or that immigration compensated for the effect of treatment. One difference between the Scottish (populations 3 and 4) and English (populations 1 and 2) populations was in changes in young-to-old ratios from December to spring (Fig. 1c). The increase in the ratios of young to old males in populations 3 and 4 reflected a replacement of old males by young that was similar in proportions on the testosterone and control areas. We suspect that this might have been due to greater predation in Scotland, where predators are more numerous than in England<sup>20</sup>. Finally, the treatment also affected female density, which was  $52 \pm 10\%$  lower on the treated than on the control areas in spring and summer (Fig. 1d; Table 1).

This experiment shows that a short-term alteration of spacing behaviour through increased testosterone levels can markedly reduce recruitment and subsequent population density, and can cause population declines. Intrinsic processes, such as changes in aggressiveness influencing recruitment, can therefore have a pivotal role in population dynamics. Over the past 40 years, territorial behaviour and parasites have been established as the two likely causal mechanisms of the red grouse population cycles<sup>6,21</sup>. Other

**Table 1** Effect of treatment on changes over time in grouse numbers

| Dependent variables  | Count data* |                           |        |                  |        |                             |       | Playback data† |                          |       |
|----------------------|-------------|---------------------------|--------|------------------|--------|-----------------------------|-------|----------------|--------------------------|-------|
|                      | d.f.        | Males per km <sup>2</sup> |        | Young/old males‡ |        | Females per km <sup>2</sup> |       | d.f.           | Males per playback point |       |
|                      |             | F                         | P      | F                | P      | F                           | P     |                | F                        | P     |
| Population           | 3,3         | 8.90                      | NS     | 3.92             | NS     | 4.12                        | NS    | 2,2            | 0.04                     | NS    |
| Treatment§           | 1,3         | 38.20                     | <0.01  | 11.53            | <0.05  | 10.27                       | <0.05 | 1,2            | 54.29                    | <0.05 |
| Period 1             | 1,12        | 49.97                     | <0.001 | 10.61            | <0.01  | 10.38                       | <0.01 | 1,8            | 14.83                    | <0.01 |
| Period 2             | 1,12        | 8.93                      | <0.05  | 7.12             | <0.05  | 0.68                        | NS    | 1,8            | 0.20                     | NS    |
| Period 1 × treatment | 1,12        | 29.26                     | <0.001 | 23.13            | <0.001 | 5.65                        | <0.05 | 1,8            | 25.33                    | <0.01 |
| Period 2 × treatment | 1,12        | 1.75                      | NS     | 1.48             | NS     | 9.47                        | <0.01 | 1,8            | 0.42                     | NS    |

\*Generalized linear mixed models include the population × treatment interaction as a random effect.

†Generalized linear mixed model includes the population × treatment, population × treatment × time, and population × treatment × playback-point interactions as random effects. F and P values are those obtained after controlling for weather scores (rain, cloud, wind) and time of day (a.m. versus p.m.) included as fixed effects.

‡Dependent variable is the number of young males, with the number of old males (log-transformed) included as an offset in the model.

§Control versus testosterone.

||The categorical effect of time was decomposed into two time periods that contrasted differences before (period 1, differences between August and December) and after treatment (period 2, differences between December and the next spring or summer). This was done by coding period 1 as 2, -1 and -1, and period 2 as 0, 1 and -1, for the August, December and spring counts, respectively.

processes, such as predation and food quality, might interact with these regulatory mechanisms to shape the observed dynamics, but do not have the potential to cause cycles<sup>6,22</sup>. The 'parasite' hypothesis states that parasite-induced reduction in grouse fecundity, coupled with low parasite aggregation in the host population, would be highly destabilizing and cause population cycles<sup>23,24</sup>. Recent experiments carried out in England supported this hypothesis, showing that a population-level reduction in *Trichostrongylus tenuis* parasites reduced population declines<sup>3,21</sup>. We have shown here, with consistent results across regions (England and Scotland), that changes in aggressiveness, lagging behind density<sup>15,16</sup>, also have a role in the unstable population dynamics. The two mechanisms are thus not necessarily region-specific, and might be operating simultaneously in cyclic grouse populations. This raises the intriguing issue of complexity. Either cycles are caused by a single dominant mechanism, which might vary geographically<sup>9</sup>, or extrinsic and intrinsic factors interact to cause the unstable dynamics<sup>1,2</sup>. Parasite transmission increases with rainfall<sup>25</sup>, so the relative dominance of these two mechanisms may vary with climate. Parasites are also likely to interact with testosterone and aggressive behaviour<sup>26,27</sup>. The crucial experiments should now test under which conditions parasites or territorial behaviour may dominate the system, and how the mechanisms may interact to generate unstable population dynamics. □

## Methods

### Experimental protocol

The study was carried out on four grouse populations, two located in northern England in autumn 2000 (populations 1 and 2; Catterick and Moorhouse), and two in northeast Scotland in autumn 2001 (populations 3 and 4; Glen Dye and Edinglassie). All study populations were cyclic and increasing when the study was conducted. On each population, two 1-km<sup>2</sup> areas separated by 0.5 km (buffer area) were selected and randomly assigned a treatment (control or testosterone). From September to November, we caught grouse by dazzling them with a lamp and netting them at night. A total of 342 males were caught, ringed and colour-marked with wing tags. We used different colours for young and old males in a given area, allowing us to age them during subsequent observations. We implanted 153 old males (an average of 92% and 86% of old males counted on the testosterone and control areas, respectively). Each was implanted with two silastic tubes (each 20 mm long, 0.62 mm inner diameter and 0.95 mm outer diameter) that were either empty (control areas) or filled with testosterone propionate (testosterone areas), and sealed with glue at both ends. Implants were inserted into the chest, between skin and breast muscles, under local anaesthesia. Testosterone implants increase aggressiveness, as measured by territorial call rate and territory size of red grouse<sup>17</sup>. During laboratory trials, the length of the tubing was determined so that the testosterone implants increased testosterone levels for three months (during the whole autumn territorial activity period). We held all the appropriate licences to conduct these procedures.

### Density estimates

We used counts with dogs to monitor changes in grouse numbers and in the ratio of young to old males. Counts were made in the morning and repeated before the experiment (August), at the end of the autumn (December), and in the following spring (April) on populations 3 and 4. In the case of populations 1 and 2, counts were made in August instead of April because foot-and-mouth disease prevented access with dogs. Adult density in late summer is highly correlated with adult density in spring<sup>22</sup>. We therefore used the number of adults counted in August as a measure of relative spring density on populations 1 and 2. In all populations, breeding success (mean number of young fledged per pair in August) before the experiment did not differ significantly between the control and testosterone areas. From the summer counts, before treatment, we estimated the number of male or female grouse available for recruitment as the number of adults plus half the number of young counted in a given area (the sex ratio at fledging is 1:1; R. Moss, unpublished observations). Young-to-old ratios before treatment were those from males caught and tagged in each area. Thereafter, we used re-sightings of tagged males to estimate young-to-old ratios.

On three of the four populations (2, 3 and 4), we also used playbacks to monitor changes in male density. At nine viewpoints regularly spaced within each area, we played a male territorial call four times. We scanned the surrounding area for 5 min and recorded all the males calling or showing within a 100 m radius of each point. We used this number of males as a measure of male density. Playbacks were carried out in the morning (8:00–12:00) or afternoon (15:00–18:00), and repeated at the same points before the experiment (early September), at the end of the autumn (December) and in spring (early May). For each scan, we scored weather conditions as follows: wind, 0 (none or light) or 1 (moderate); rain, 0 (no) or 1 (light); and cloud cover, 0 (clear), 1 (patchy) or 2 (cloudy).

### Statistical analyses

We used generalized linear mixed models (Glimmix procedure, SAS 8.01, SAS Institute) to analyse the count and playback data. Models were fitted to the data using a Poisson error

distribution corrected for over-dispersion. Population (1–4), time (decomposed into two contrasts), treatment (control versus testosterone) and time × treatment interactions were included as fixed effects. For count data, models included the population × treatment interaction as a random effect to ensure that the treatment effect was tested against the appropriate error variance. For playback data, the model also included the population × treatment × time, and the population × treatment × playback-point interactions as random effects. The former ensures that the time contrasts are tested against the correct error variance, and the latter accounts for consistent effects of each playback point. Time of day (a.m. versus p.m.) and weather conditions might affect males' response to playbacks and were controlled for in the analysis of the playback data.

Received 24 October; accepted 31 December 2002; doi:10.1038/nature01395.

- Krebs, C. J. *et al.* Impact of food and predation on the snowshoe hare cycle. *Science* **269**, 1112–1115 (1995).
- Stenseth, N. C., Bjørnstad, O. N. & Falck, W. Is spacing behaviour coupled with predation causing the microtine density cycle? A synthesis of current process-oriented and pattern-oriented studies. *Proc. R. Soc. Lond. B* **263**, 1423–1435 (1996).
- Hudson, P. J., Dobson, A. P. & Newborn, D. Prevention of population cycles by parasite removal. *Science* **282**, 2256–2258 (1998).
- Turchin, P., Taylor, A. D. & Reeve, J. D. Dynamical role of predators in population cycles of a forest insect: An experimental test. *Science* **285**, 1068–1071 (1999).
- Grenfell, B. T., Bjørnstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
- Moss, R. & Watson, A. Population cycles in birds of the grouse family (Tetraonidae). *Adv. Ecol. Res.* **32**, 53–111 (2001).
- Korpimäki, E., Norrdahl, K., Klemola, T., Pettersen, T. & Stenseth, N. C. Dynamic effects of predators on cyclic voles: Field experimentation and model extrapolation. *Proc. R. Soc. Lond. B* **269**, 991–997 (2002).
- Ostfeld, R. S., Canham, C. D. & Pugh, S. R. Intrinsic density-dependent regulation of vole populations. *Nature* **366**, 259–261 (1993).
- Berryman, A. A. in *Population Cycles: The Case for Trophic Interactions* (ed. Berryman, A. A.) 3–28 (Oxford Univ. Press, Oxford, UK, 2002).
- Lambin, X., Krebs, C. J., Moss, R. & Yoccoz, N. G. in *Population Cycles: The Case for Trophic Interactions* (ed. Berryman, A. A.) 155–176 (Oxford Univ. Press, Oxford, UK, 2002).
- Lambin, X. & Yoccoz, N. G. The impact of population kin-structure on nestling survival in Townsend's voles, *Microtus townsendii*. *J. Anim. Ecol.* **67**, 1–16 (1998).
- Bjørnstad, O. N. & Grenfell, B. T. Noisy clockwork: Time series analysis of population fluctuations in animals. *Science* **293**, 638–643 (2001).
- Blasius, B., Huppert, A. & Stone, L. Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature* **399**, 354–359 (1999).
- Mountford, M. C., Watson, A., Moss, R. & Parr, R. in *Red Grouse Population Processes* (eds Lance, A. N. & Lawton, J. H.) 78–83 (British Ecological Society, Ascot, UK, 1990).
- Watson, A., Moss, R., Parr, R., Mountford, M. D. & Rothery, P. Kin landownership, differential aggression between kin and non-kin, and population fluctuations in red grouse. *J. Anim. Ecol.* **63**, 39–50 (1994).
- Moss, R., Watson, A. & Parr, R. Experimental prevention of a population cycle in red grouse. *Ecology* **77**, 1512–1530 (1996).
- Moss, R., Parr, R. & Lambin, X. Effects of testosterone on breeding density, breeding success and survival of red grouse. *Proc. R. Soc. Lond. B* **258**, 175–180 (1994).
- Charnov, E. L. & Finnerty, J. P. Vole population cycles: A case for kin selection. *Oecologia* **45**, 1–2 (1980).
- Matthiopoulos, J., Moss, R. & Lambin, X. The kin-facilitation hypothesis for red grouse population cycles: Territory sharing between relatives. *Ecol. Model.* **127**, 53–63 (2000).
- Hudson, P. J., Newborn, D. & Robertson, P. J. Seasonal and geographical patterns of mortality in red grouse populations. *Wildl. Biol.* **2**, 79–88 (1997).
- Hudson, P. J. *et al.* Trophic interactions and population growth rates: describing patterns and identifying mechanisms. *Phil. Trans. R. Soc. Lond. B* **357**, 1259–1271 (2002).
- Thirgood, S. J., Redpath, S. M., Rothery, P. & Aebischer, N. J. Raptor predation and population limitation in red grouse. *J. Anim. Ecol.* **69**, 504–516 (2000).
- Dobson, A. P. & Hudson, P. J. Regulation and stability of a free-living host-parasite system—*Trichostrongylus tenuis* in red grouse. II. Population models. *J. Anim. Ecol.* **61**, 487–498 (1992).
- Hudson, P. J., Newborn, D. & Dobson, A. P. Regulation and stability of a free-living host-parasite system—*Trichostrongylus tenuis* in red grouse. I. Monitoring and parasite reduction experiments. *J. Anim. Ecol.* **61**, 477–486 (1992).
- Hudson, P. J., Dobson, A. P. & Newborn, D. in *Ecology and Genetics Of Host-Parasite Interactions* (eds Rollinson, D. & Andersson, R. M.) 77–89 (Academic, London, UK, 1985).
- Fox, A. & Hudson, P. J. Parasites reduce territorial behaviour in red grouse (*Lagopus lagopus scoticus*). *Ecol. Lett.* **4**, 139–143 (2001).
- Folstad, I. & Karter, A. J. Parasites, bright males, and the immunocompetence handicap. *Am. Nat.* **139**, 603–622 (1992).

**Acknowledgements** We thank the British army, English Nature, and the Dalhousie and Dunecht estates for allowing us access to their moors. We particularly thank T. P. J. Helps, C. McCarthy, J. Adamson, D. Calder and A. Dykes for their help in organizing the fieldwork. We also thank R. Cox, N. Green, D. Luccini and J. Irvine for their help with the fieldwork. D. Elston helped with the statistical analyses. S. Albon, M. Harris, X. Lambin, M. Marquiss, R. Moss and R. Van Der Wal made helpful comments on previous drafts of the manuscript. This work was funded by the Natural Environment Research Council.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to F.M. (e-mail: frm@ceh.ac.uk).



# Auxin promotes *Arabidopsis* root growth by modulating gibberellin response

Xiangdong Fu & Nicholas P. Harberd

John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK

The growth of plant organs is influenced by a stream of the phytohormone auxin that flows from the shoot apex to the tip of the root<sup>1</sup>. However, until now it has not been known how auxin regulates the cell proliferation and enlargement that characterizes organ growth. Here we show that auxin controls the growth of roots by modulating cellular responses to the phytohormone gibberellin (GA). GA promotes the growth of plants by opposing the effects of nuclear DELLA protein growth repressors<sup>2–8</sup>, one of which is *Arabidopsis* RGA (for repressor of *ga1-3*)<sup>9,10</sup>. GA opposes the action of several DELLA proteins by destabilizing them, reducing both the concentration of detectable DELLA proteins and their growth-restraining effects<sup>9–14</sup>. We also show that auxin is necessary for GA-mediated control of root growth, and that attenuation of auxin transport or signalling delays the GA-induced disappearance of RGA from root cell nuclei. Our observations indicate that the shoot apex exerts long-distance control on the growth of plant organs through the effect of auxin on GA-mediated DELLA protein destabilization.

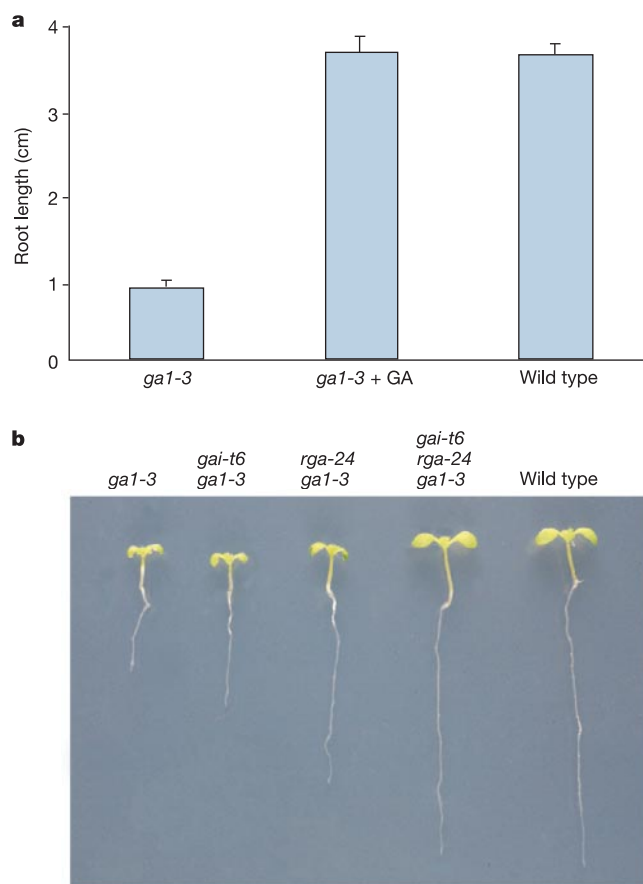
Auxin mediates at least some of its effects on plant growth by inducing the proteasome-mediated degradation of AUX/IAA proteins<sup>15,16</sup>. However, the mechanism by which auxin regulates growth remains unclear. Furthermore, it is not known how the auxin and GA signalling pathways interact with one another<sup>17</sup>. Here we demonstrate a molecular link between auxin-mediated and GA-mediated growth regulation: auxin controls the growth of *Arabidopsis* roots through the modulation of the cellular response to GA.

Although GA has long been known to regulate shoot growth, its role in the regulation of root growth was less clear. We showed that primary roots of the GA-deficient *Arabidopsis* *ga1-3* mutant<sup>5</sup> are shorter than those of the wild type, whereas GA-treated *ga1-3* roots are of a similar length to those of the wild type (Fig. 1a). Thus, GA regulates root growth. *Arabidopsis* contains genes encoding five DELLA proteins (GAI, RGA, RGL1, RGL2 and RGL3), and genetic analysis has shown that some of these proteins mediate specific components of the GA response<sup>2,3,5–8</sup>. We showed that roots lacking GAI (*gai-t6 ga1-3*) or RGA (*rga-24 ga1-3*) are longer than *ga1-3* controls (Fig. 1b), whereas a lack of both GAI and RGA (*gai-t6 rga-24 ga1-3*) substantially suppresses the *ga1-3* phenotype (Fig. 1b). Thus, GAI and RGA act together in the regulation of root (and shoot<sup>5,6</sup>) growth, as growth repressors whose function is opposed by GA.

In 1958, Brian and Hemming showed that the stem internodes of GA-deficient dwarf mutant pea plants failed to elongate in response to GA treatments if the shoot apex was removed, and that an intact shoot apex is necessary for a normal response to GA<sup>18</sup>. We tested the role of the shoot apex in regulating GA-responsive root growth in *Arabidopsis*. Removal of the shoot apex (decapitation) inhibited the growth of primary roots of wild-type *Arabidopsis* seedlings (Fig. 2a). Application of auxin (IAA) to the site from which the shoot apex had been removed restored the growth of the roots of decapitated seedlings to normal (Fig. 2a), which is consistent with previous observations that externally applied auxin can stimulate root growth<sup>19</sup>. Thus, shoot apex-derived auxin passes to the root via

the polar auxin transport stream and promotes root growth. Next we found that the roots of decapitated *ga1-3* plants were much less responsive to GA than were the roots of intact *ga1-3* plants (Fig. 2b). In addition, application of IAA to the site from which the shoot apex had been removed resulted in recovery of the response of roots of *ga1-3* seedlings to GA (Fig. 2b). Thus, shoot apex-derived auxin promotes root growth by facilitating the response of root cells to GA.

We next examined the effect of removing the shoot apex of *pRGA:GFP-RGA* plants on the GA-induced disappearance from root cell nuclei of the fusion protein of green fluorescent protein (GFP) with RGA (by the detection of GFP fluorescence using confocal microscopy; see Methods). As shown previously<sup>9,10</sup>, treatment of (intact) *pRGA:GFP-RGA* plants with GA resulted in a rapid disappearance of GFP-RGA from root nuclei, in both the tip and elongating regions of the root (Fig. 2c). However, after 4 h of treatment with GA, GFP-RGA was still detectable in root nuclei of decapitated seedlings (although fainter than in the roots of untreated control and decapitated seedlings; Fig. 2c). Application of IAA to the site from which the shoot apex had been removed restored the rapidity of the GA-induced disappearance of GFP-RGA from root nuclei to a similar rate to that seen in intact control seedlings (Fig. 2c). Thus, GA-induced disappearance of GFP-RGA is delayed in the absence of shoot apex-derived auxin.



**Figure 1** GA regulates root elongation via DELLA proteins. **a**, Mean length ( $\pm$ s.e.;  $n = 30$ ) of 5-day-old primary seedling roots of *ga1-3*, *ga1-3* plants grown on  $1 \mu\text{M}$  GA<sub>3</sub> (*ga1-3* + GA) and wild-type (Landsberg *erecta*) plants. **b**, Representative 5-day-old primary seedling roots of selected genotypes. Mean lengths were determined ( $\pm$ s.e.,  $n = 30$ ): *ga1-3*,  $1.0 \pm 0.1$  cm; *gai-t6 ga1-3*,  $1.3 \pm 0.2$  cm; *rga-24 ga1-3*,  $2.0 \pm 0.1$  cm; *gai-t6 rga-24 ga1-3*,  $3.6 \pm 0.2$  cm; wild type,  $3.7 \pm 0.2$  cm.

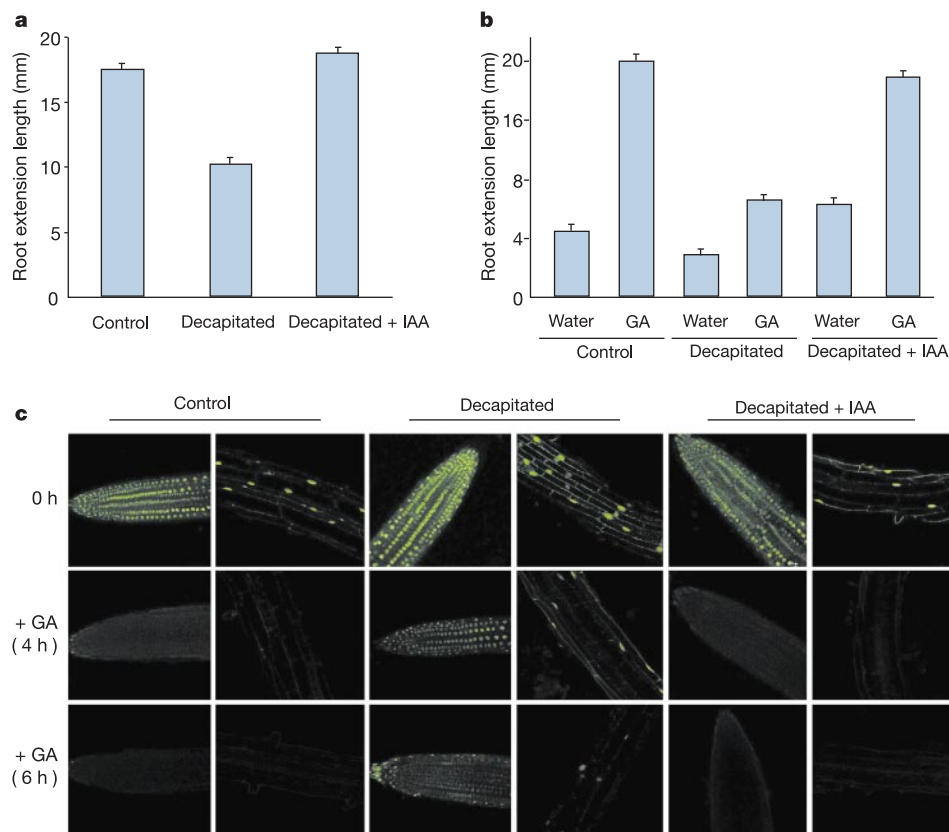
These results indicate that, in wild-type plants, auxin transported in a polar manner promotes a GA-induced reduction in RGA concentration, thus releasing root growth from RGA-mediated restraint.

In further experiments we found that 1-*N*-naphthylphthalamic acid (NPA; a specific inhibitor of auxin efflux<sup>20</sup>) prevents exogenous GA from restoring normal growth to GA-deficient (*gai-3*) roots (Fig. 3a). Furthermore, the growth of the roots of plants lacking GAI and RGA was relatively resistant to the effects of NPA (Fig. 3a). For example, whereas the growth of GA-treated *gai-3* roots is inhibited by NPA, the effect of NPA on the growth of GA-treated *gai-t6 rga-24* roots is much less (Fig. 3a). In addition, the growth of *gai-t6 rga-24* roots is more resistant to the effects of NPA than is that of wild-type roots (data not shown). These results indicate that NPA exerts its effects on root growth through GAI and RGA. We also found that NPA altered the kinetics of the GA-induced disappearance of GFP-RGA (Fig. 3b). After a 2-h GA treatment of *pRGA:GFP-RGA* roots pretreated with NPA, GFP-RGA was still clearly detectable in root nuclei, although the fluorescence was fainter than in the control (no GA) root tips (Fig. 3b). After 4 h of GA treatment, nuclear GFP-RGA had largely disappeared from the tip but remained clearly detectable in the cellular elongation and hair-forming zones of the root (Fig. 3b). These results indicate that NPA causes GFP-RGA to be relatively resistant to GA-mediated destabilization.

Auxin transport is thought to be mediated by specific membrane-associated cellular influx<sup>21</sup> and efflux<sup>22</sup> regulators. In particular, polar auxin transport through stem and root vasculature is con-

sidered to be largely controlled by the efflux regulator AtPIN1, a protein located at the basal end of auxin-transporting cells<sup>23</sup>, and is thought to be the site of NPA-mediated inhibition of polar auxin transport<sup>24</sup>. We tested the possibility that AtPIN1 function is necessary for normal GA-mediated destabilization of GFP-RGA, using a method for the transient induction of systemic gene silencing based on RNA-mediated interference (RNAi)<sup>25,26</sup>. We reasoned that, after the *Agrobacterium*-mediated introduction of a gene-silencing (RNAi) DNA construct into *Arabidopsis* root cells, silencing should spread systemically from infected cells to the remaining cells of the root. We then showed that the infection of *pRGA:GFP-RGA* roots with *Agrobacterium* containing an anti-AtPIN1 RNAi construct caused the rapid disappearance of AtPIN1 mRNA detectable by polymerase chain reaction (PCR) with reverse transcription (Fig. 4a). In addition, preinfection of *pRGA:GFP-RGA* roots with *Agrobacterium* containing the anti-AtPIN1 RNAi construct caused GFP-RGA to be relatively resistant to treatments with GA. As shown in Fig. 4b, the roots of *pRGA:GFP-RGA* plants preinfected with *Agrobacterium* containing the anti-AtPIN1 RNAi construct retained GFP-RGA for up to 4 h after the onset of treatments with GA, whereas plants preinfected with *Agrobacterium* containing vector-only control constructs did not. These observations indicate that AtPIN1 function promotes the GA-mediated destabilization of GFP-RGA.

Mutations at the *AXR1* locus reduce auxin response through attenuation of the function of the SCF<sup>TIR1</sup> E3 polyubiquitin ligase<sup>15,27</sup>. We tested the effect of the *axr1-12* (ref. 15) mutation on the kinetics of the GA-induced disappearance of GFP-RGA, and

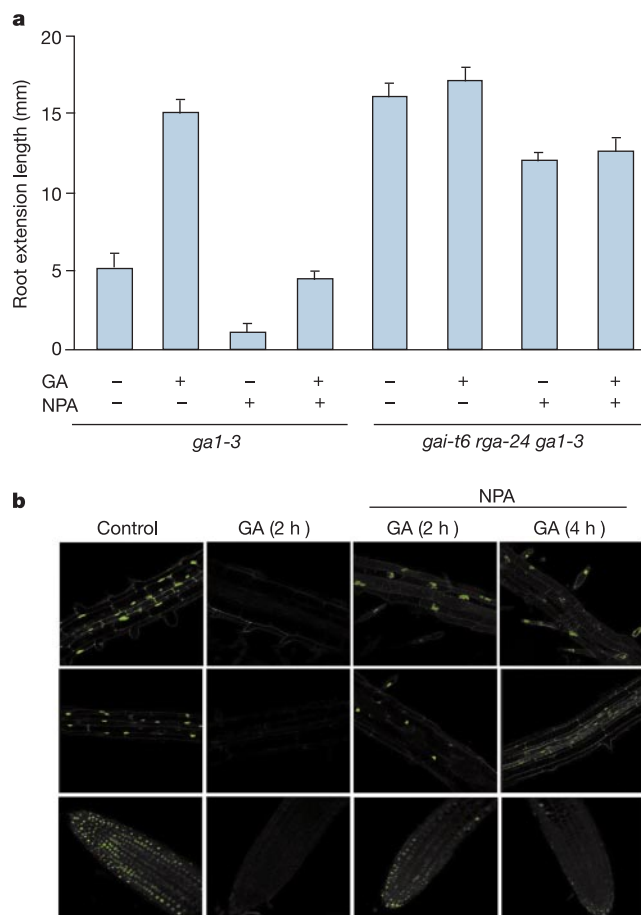


**Figure 2** Shoot apex-derived auxin controls root growth. **a**, Wild-type root growth is stimulated by polar auxin transport. Extent of root growth over a 3-d period (see Methods) is shown as mean ( $\pm$ s.e.;  $n = 60$ ) for control seedlings, decapitated seedlings and seedlings in which IAA was applied to the site from which the shoot apex had been

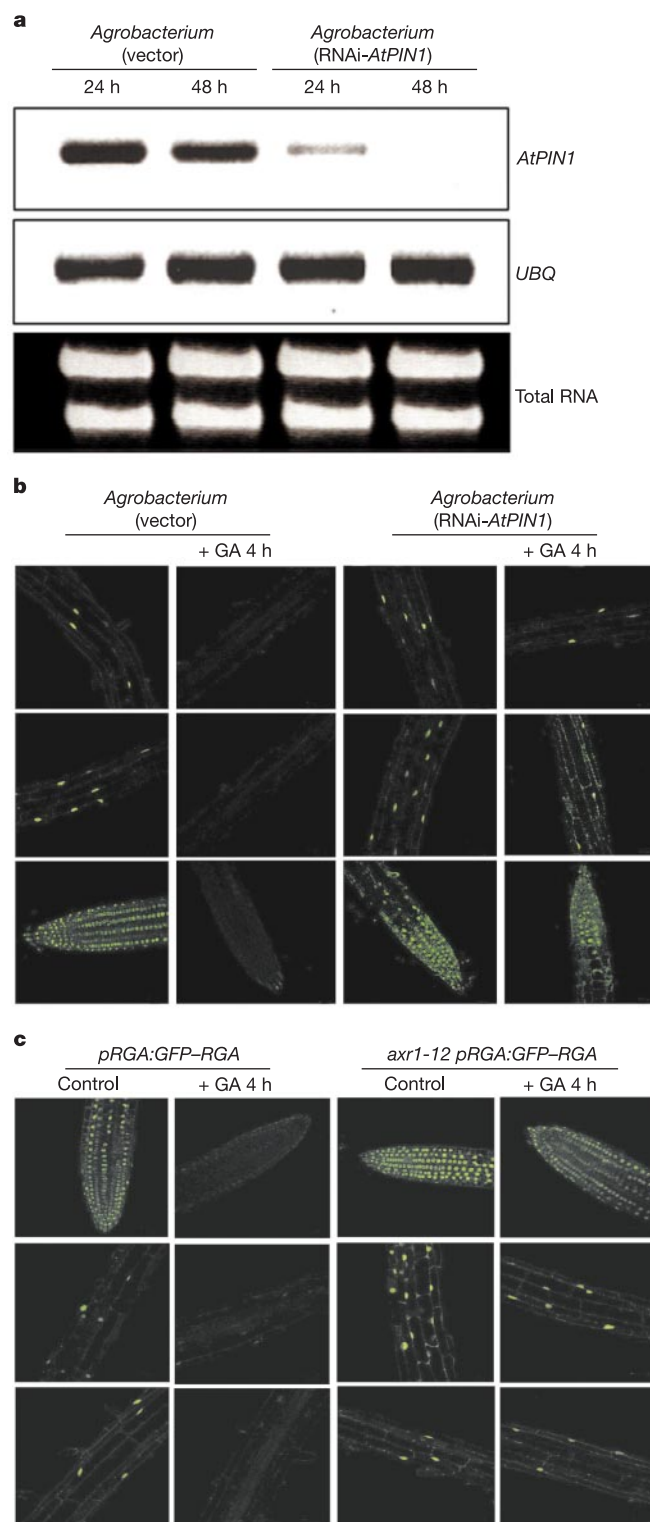
removed. **b**, GA regulates *gai-3* root growth in the presence of shoot apex-derived auxin. Extent of growth was determined as in **a** except that  $n = 30$ . **c**, GFP fluorescence of GA-treated ( $1 \mu\text{M}$  GA<sub>3</sub>) *pRGA:GFP-RGA* primary seedling roots. Double columns show root tip (left) and elongation zone (right).

found that GFP-RGA was detectable in root nuclei for a longer period after the initiation of treatment with GA in an *axr1-12* background than it was in an *AXR1* background (Fig. 4c). Thus, in agreement with our previous observations, a mutation affecting auxin signalling compromises the GA-dependent reduction in RGA level. However, it is clear that *axr1* mutations can affect the activity of signalling pathways that are additional to the auxin signalling pathway<sup>28</sup>, and it is possible that these additional pathways also affect GFP-RGA levels.

We have shown that shoot apex-derived auxin regulates root growth by modulating the effect of GA on the levels of a DELLA growth-repressing protein. When polar auxin transport or auxin signalling are attenuated, the GA-induced disappearance of nuclear RGA is delayed and growth is restrained. These observations indicate that auxin promotes the growth of roots by enhancing the GA-induced destabilization of RGA, and perhaps of other DELLA proteins. Although our results relate to the growth of roots, it is likely that they are indicative of the control of growth in other plant organs. For example, our findings provide a ready explanation for the observation that the internodes of decapitated GA-deficient pea plants are less responsive to GA than the internodes of intact GA-deficient pea plants<sup>18</sup>. Interestingly, it has recently been shown that the level of activity of enzymes involved in the biosynthesis of GA is also dependent on the polar auxin transport stream<sup>29</sup>. Thus, both GA biosynthesis and the GA



**Figure 3** NPA inhibits the effect of GA on root growth and disappearance of GFP-RGA from root cell nuclei. **a**, Mean growth ( $\pm$ s.e.;  $n = 40$ ), over 3 d, of *ga1-3* and *gai-t6 rga-24 ga1-3* primary seedling roots treated with GA and/or NPA. **b**, GFP fluorescence in the tip (bottom), elongation zone (middle) and epidermal hair zone (top) of *pRGA:GFP-RGA* primary roots (for details of treatments see Methods).



**Figure 4** Decrease in *AtPIN1* or *AXR1* function delays GA-induced disappearance of GFP-RGA. **a**, Co-cultivation with *Agrobacterium* containing RNAi-*AtPIN1* (but not with *Agrobacterium* containing vector only) causes a decrease in *AtPIN1* but not in control (ubiquitin; *UBQ*) transcript levels. **b**, Silencing of *AtPIN1* affects GA-induced disappearance of GFP-RGA. GFP fluorescence in the tip (bottom), elongation zone (middle) and epidermal hair zone (top) of *Agrobacterium*-infected *pRGA:GFP-RGA* primary roots after treatment with GA (100  $\mu$ M  $GA_3$ ) is shown. **c**, The *axr1-12* mutation affects the GA-induced disappearance of GFP-RGA. GFP fluorescence in *pRGA:GFP-RGA* and *axr1-12 pRGA:GFP-RGA* primary roots (zones as above) after treatment with GA (100  $\mu$ M  $GA_3$ ) is shown.



response pathway are subject to regulation by auxin, indicating that auxin has several inputs into the regulation of growth through its effects on DELLA protein activity. Of course, it is also possible that auxin additionally regulates growth through one or more pathways that are independent of DELLA protein.

Although it has long been known that the various phytohormones (such as auxin, GA, ethylene, cytokinin, abscisic acid and brassinosteroid) interact in the regulation of plant growth and development, the mechanisms of these interactions have remained poorly understood<sup>17</sup>. Here we have demonstrated that the properties of a signalling component originally identified as being part of the GA-response signalling pathway are modulated by a different phytohormone, namely auxin. Growth is a quantitative phenomenon, and our findings identify differential auxin flux, with consequent differential effects on the stability of DELLA proteins, as one possible quantitative modulator of plant growth. □

## Methods

### Materials

Wild-type *Arabidopsis thaliana* (Landsberg erecta), *gal1-3*, *gai-16*, *rga-24*, *axr1-12* and *pRGA:GFP-RGA*, and lines containing combinations thereof, were as described previously<sup>5,6,9,10,15</sup>. Previous studies have shown that the GFP-RGA fusion protein encoded by *pRGA:GFP-RGA* complements a *rga* loss-of-function mutation and that the GA-induced disappearance of GFP-RGA (as determined by loss of detectable GFP fluorescence) is correlated with the disappearance of immunologically detectable GFP-RGA<sup>9,10</sup>. The *axr1-12 pRGA:GFP-RGA* line was isolated from the F<sub>3</sub> of a cross between *axr1-12* (Columbia background) and *pRGA:GFP-RGA*. The *AXR1 pRGA:GFP-RGA* line (shown in Fig. 4c) was another segregant from this cross.

### Root growth experiments

All seeds were pretreated at 4 °C with 1 μM gibberellic acid (GA<sub>3</sub>) for 5 d to synchronize germination, then thoroughly washed. In the experiments shown in Fig. 1, seeds were then placed on GM<sup>3</sup> agar plates with or without 1 μM GA<sub>3</sub> (GA) and stacked vertically in a growth room (22 °C; 16-h photoperiod). Root length was measured from root tip to base of hypocotyl. For experiments involving decapitation (Fig. 2a, b), 5-day-old seedlings (grown on GM agar as above) were decapitated (or not) by removal of the shoot apex. Decapitated plants were then maintained for 3 d on GM medium to confirm that the plants were still growing and that the apex had been fully removed. Root growth in response to a subsequent 3-d treatment (at the site from which the apex had been removed) with water or 1 ng l<sup>-1</sup> indoleacetic acid (IAA) applied daily as 0.1-μl droplets was measured by marking and recording root tip positions at the beginning and end of the treatment. We also found that the growth of decapitated seedlings on medium containing 1 ng l<sup>-1</sup> IAA promoted root growth (data not shown). Five-day-old *gal1-3* seedlings were similarly decapitated and maintained for 3 d or left intact, and were then grown on 1 μM GA<sub>3</sub> (GA) or control (water) (results shown in Fig. 2b). In further experiments 5-day-old seedlings were pretreated with or without NPA (25 μM) for 3 d and then moved to agar medium containing (or lacking) 1 μM GA<sub>3</sub> (GA) and/or NPA (results shown in Fig. 3a).

### Experiments involving detection of GFP-RGA

Seedlings containing *pRGA:GFP-RGA* were treated for 0–6 h with 1 or 100 μM GA<sub>3</sub> (refs 9, 10) (we found that GFP-RGA disappears more slowly in the presence of 1 μM GA<sub>3</sub> than in the presence of 100 μM GA<sub>3</sub>, but that in intact seedlings fluorescence became undetectable after 4 h of either treatment). Confocal microscope images of primary roots were from a Leica inverted confocal laser microscope with ×40 objectives, at an excitation wavelength of 488 nm. Within each figure, images were obtained with a constant set of microscopic and image intensity parameters. Images (Fig. 2c) were from roots of decapitated seedlings 3 d after decapitation (with non-decapitated controls); other images are from seedlings 3 d after decapitation with subsequent daily application of 0.1 μl of 1 ng l<sup>-1</sup> IAA to the site from which the apex had been removed. Seedlings in Fig. 2c were treated with 1 μM GA<sub>3</sub> for the durations shown. For the experiments shown in Fig. 3b, 5-day-old seedlings were pretreated with 25 μM NPA or water (control) for 3 d, then treated with GA<sub>3</sub> (100 μM) plus NPA as appropriate (or water) for the durations shown. For the experiments shown in Fig. 4b, 5-day-old seedlings were co-cultivated for 48 h with *Agrobacterium* containing (RNAi-*AtPIN1*) or lacking the *AtPIN1* silencing construct (vector) and were subsequently treated for 4 h with GA (100 μM GA<sub>3</sub>) or water. For the experiments shown in Fig. 4c, plants were treated with 100 μM GA.

### Transient gene silencing

The *AtPIN1* coding sequence (a 401-base-pair fragment) was amplified by PCR with primers 5'-GGATCCTCGAGTCTTACACACAGACCAATGC-3' and 5'-ATCGATGGTACCTGGAACTGCTTGAGATCA-3', and cloned twice (fragment-intron-fragment in reverse orientation) into a CaMV35S promoter-driven expression vector to make dsRNAi-*AtPIN1*. Five-day-old *pRGA:GFP-RGA* seedlings were co-cultivated with *Agrobacterium* (strain GV3101; pretreated with 150 μM acetosyringone) containing dsRNAi-*AtPIN1* or empty vector (control) for the durations shown in Fig. 4a,

after which RNA was extracted from roots. Generation of complementary DNA, amplification by PCR (20 cycles) and blot analyses were as described previously<sup>30</sup>. The primer pair used for PCR amplification (5'-CCAAGCAGTTTCCAGACACA-3' and 5'-TCCGTTTCCGTCCTTGTCTT-3') amplifies a 608-base-pair fragment from the *AtPIN1* open reading frame.

Received 7 October; accepted 10 December 2002; doi:10.1038/nature01387.

- Lomax, T. L., Munday, G. K. & Rubery, P. H. In *Plant Hormones: Physiology, Biochemistry and Molecular Biology* (ed. Davies, P. J.) 509–530 (Kluwer, Dordrecht, 1995).
- Richards, D. E., King, K. E., Ait-ali, T. & Harberd, N. P. How gibberellin regulates plant growth and development: A molecular genetic analysis of gibberellin signalling. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 67–88 (2001).
- Peng, J. et al. The *Arabidopsis* *GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes Dev.* **11**, 3194–3205 (1997).
- Peng, J. et al. 'Green Revolution' genes encode mutant gibberellin response modulators. *Nature* **400**, 256–261 (1999).
- King, K. E., Moritz, T. & Harberd, N. P. Gibberellins are not required for stem growth in *Arabidopsis thaliana* in the absence of *GAI* and *RGA*. *Genetics* **159**, 767–776 (2001).
- Dill, A. & Sun, T.-p. Synergistic derepression of gibberellin signaling by removing *RGA* and *GAI* function in *Arabidopsis thaliana*. *Genetics* **159**, 777–785 (2001).
- Lee, S. et al. Gibberellin regulates *Arabidopsis* seed germination via *RGL2*, a *GAI/RGA*-like gene whose expression is up-regulated following imbibition. *Genes Dev.* **16**, 646–658 (2002).
- Wen, C.-K. & Chang, C. *Arabidopsis* *RGL1* encodes a negative regulator of gibberellin responses. *Plant Cell* **14**, 87–100 (2002).
- Silverstone, A. L. et al. Repressing a repressor: gibberellin-induced rapid reduction of the *RGA* protein in *Arabidopsis*. *Plant Cell* **13**, 1555–1565 (2001).
- Dill, A., Jung, H.-S. & Sun, T.-p. The DELLA motif is essential for gibberellin-induced degradation of *RGA*. *Proc. Natl Acad. Sci. USA* **98**, 14162–14167 (2001).
- Itoh, H., Ueguchi-Tanaka, M., Sato, Y., Ashikari, M. & Matsuoka, M. The gibberellin signaling pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei. *Plant Cell* **14**, 1–16 (2002).
- Chandler, P. M., Marion-Poll, A., Ellis, M. & Gubler, F. Mutants at the *Slender1* locus of barley cv Himalaya. Molecular and physiological characterization. *Plant Physiol.* **129**, 181–190 (2002).
- Gubler, F., Chandler, P. M., White, R. G., Llewellyn, D. J. & Jacobsen, J. V. Gibberellin signaling in barley aleurone cells. Control of *SLN1* and *GAMYB* expression. *Plant Physiol.* **129**, 191–200 (2002).
- Fu, X. et al. Gibberellin-mediated proteasome-dependent degradation of the barley DELLA protein *SLN1* repressor. *Plant Cell* **14**, 3191–3200 (2002).
- Gray, W. M., Kepinski, S., Rouse, D., Leyser, O. & Estelle, M. Auxin regulates SCF<sup>TIR1</sup>-dependent degradation of AUX/IAA proteins. *Nature* **414**, 271–276 (2001).
- Rogg, L. E. & Bartel, B. Auxin signalling: derepression through regulated proteolysis. *Dev. Cell* **1**, 595–604 (2001).
- Davies, P. J. (ed.) *Plant Hormones: Physiology, Biochemistry and Molecular Biology* (Kluwer, Dordrecht, 1995).
- Brian, P. W. & Hemming, H. G. Complementary action of gibberellic acid and auxins in pea internode extension. *Ann. Bot.* **22**, 1–17 (1958).
- Evans, M. L., Ishikawa, H. & Estelle, M. A. Responses of *Arabidopsis* roots to auxin studied with high temporal resolution: comparison of wild type and auxin-response mutants. *Planta* **194**, 215–222 (1994).
- Ruegger, M. et al. Reduced naphthylphthalamic acid binding in the *tir3* mutant of *Arabidopsis* is associated with a reduction in polar auxin transport and diverse morphological defects. *Plant Cell* **9**, 745–773 (1997).
- Bennett, M. J. et al. *Arabidopsis* *AUX1* gene: a permease-like regulator of root gravitropism. *Science* **273**, 948–950 (1996).
- Muller, A. et al. *AtPIN2* defines a locus of *Arabidopsis* for root gravitropism control. *EMBO J.* **17**, 6903–6911 (1998).
- Galweiler, L. et al. Regulation of polar auxin transport by *AtPIN1* in *Arabidopsis* vascular tissue. *Science* **282**, 2226–2230 (1998).
- Geldner, N. et al. Auxin transport inhibitors block *PIN1* cycling and vesicle trafficking. *Nature* **413**, 425–428 (2001).
- Baulcombe, D. RNA silencing. *Curr. Biol.* **12**, R82–R84 (2002).
- Mlotshwa, S. et al. RNA silencing and the mobile silencing signal. *Plant Cell* **14**, S289–S301 (1994).
- del Pozzo, J. C. et al. *AXR1-ECR1*-dependent conjugation of *RUB1* to the *Arabidopsis* cullin *AtCUL1* is required for auxin response. *Plant Cell* **14**, 421–433 (2002).
- Xu, L. et al. The SCF<sup>COI1</sup> ubiquitin-ligase complexes are required for jasmonate response in *Arabidopsis*. *Plant Cell* **14**, 1919–1935 (2002).
- Ross, J. J., O'Neill, D. P., Smith, J. J., Kerckhoffs, L. H. J. & Elliott, R. C. Evidence that auxin promotes gibberellin A<sub>1</sub> biosynthesis in pea. *Plant J.* **21**, 547–552 (2000).
- Fu, X. et al. Expression of *Arabidopsis* *GAI* in transgenic rice represses multiple gibberellin responses. *Plant Cell* **13**, 1791–1802 (2001).

**Acknowledgements** We thank T.-p. Sun (Duke University) for the *pRGA:GFP-RGA* line, O. Leyser for *axr1-12*, G. Calder for assistance with confocal microscopy, and P. Achard, T. Ait-ali, E. Coen, O. Leyser, D. Richards and X. Zhang for comments on the manuscript. Funding was received from the BBSRC (Core Strategic Grant to the John Innes Centre and a response mode grant).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to N.H. (e-mail: nicholas.harberd@bbsrc.ac.uk).

# Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain

Daniel J. Cua\*, Jonathan Sherlock\*, Yi Chen\*, Craig A. Murphy\*, Barbara Joyce\*, Brian Seymour\*, Linda Lucian\*, Wayne To†, Sylvia Kwan†, Tatyana Churakova†, Sandra Zurawski†, Maria Wiekowski‡, Sergio A. Lira§, Daniel Gorman||, Robert A. Kastelein|| & Jonathon D. Sedgwick\*

\* Departments of Immunology, || Department of Genomics, and † Department of Protein and Antibody Technology, DNAX Research Inc., Palo Alto, California 94304-1104, USA

‡ Department of Immunology, Schering-Plough Research Institute, Kenilworth, New Jersey 07033, USA

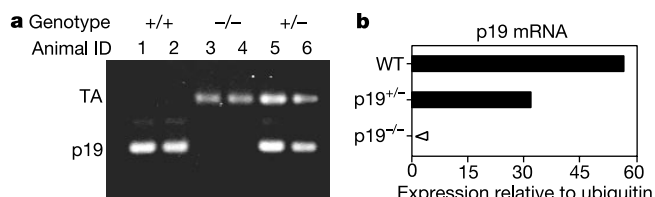
§ Present address: Immunobiology Center, Mount Sinai School of Medicine, New York, New York 10029-6574, USA

Interleukin-12 (IL-12) is a heterodimeric molecule composed of p35 and p40 subunits. Analyses *in vitro* have defined IL-12 as an important factor for the differentiation of naive T cells into T-helper type 1 CD4<sup>+</sup> lymphocytes secreting interferon- $\gamma$  (refs 1, 2). Similarly, numerous studies<sup>3–7</sup> have concluded that IL-12 is essential for T-cell-dependent immune and inflammatory responses *in vivo*, primarily through the use of IL-12 p40 gene-targeted mice and neutralizing antibodies against p40. The cytokine IL-23, which comprises the p40 subunit of IL-12 but a different p19 subunit<sup>8</sup>, is produced predominantly by macrophages and dendritic cells, and shows activity on memory T cells. Evidence from studies of IL-23 receptor expression<sup>9</sup> and IL-23 overexpression in transgenic mice<sup>10</sup> suggest, however, that IL-23 may also affect macrophage function directly. Here we show, by using gene-targeted mice lacking only IL-23 and cytokine replacement studies, that the perceived central role for IL-12 in autoimmune inflammation, specifically in the brain, has been misinterpreted and that IL-23, and not IL-12, is the critical factor in this response. In addition, we show that IL-23, unlike IL-12, acts more broadly as an end-stage effector cytokine through direct actions on macrophages.

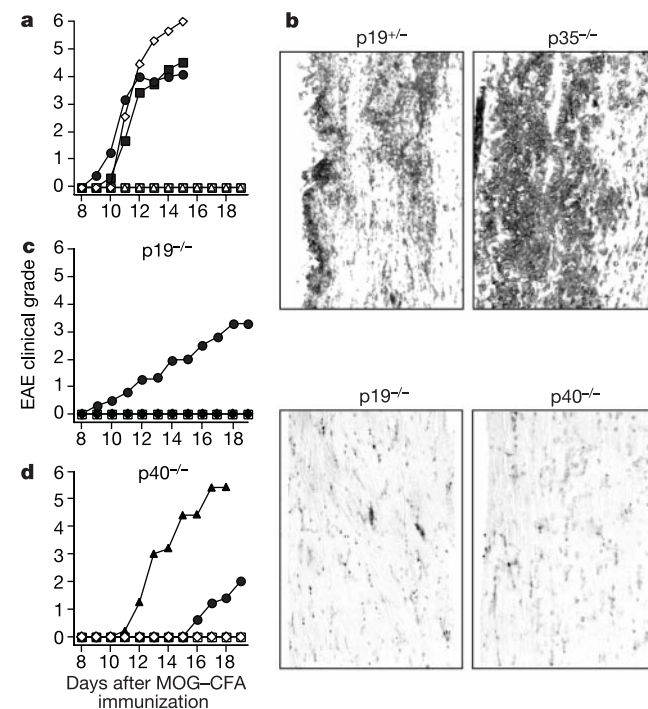
To determine the relative contributions of IL-12 and IL-23 in autoimmune inflammation, we compared the susceptibility of mice lacking only IL-23 (*p19*<sup>−/−</sup>), only IL-12 (*p35*<sup>−/−</sup>) or both cytokines (*p40*<sup>−/−</sup>) to experimental autoimmune encephalomyelitis (EAE), a autoimmune inflammatory disease of the central nervous system (CNS) that is mediated by T-helper type 1 (T<sub>H</sub>1) cells and inflammatory macrophages. The *IL-23 p19*<sup>−/−</sup> mice used in this

study were generated by removing the whole *p19* locus (see Supplementary Fig. 1). Polymerase chain reaction (PCR) analysis of genomic DNA showed the presence of the targeted allele but not the wild-type *IL-23* allele in *p19*<sup>−/−</sup> mice (Fig. 1a). In addition, quantitative analysis of lymphoid tissues by PCR with reverse transcription (RT-PCR) showed the presence of *p19* mRNA in wild-type and *p19*<sup>+/−</sup> mice but not in *p19*<sup>−/−</sup> mice, indicating successful ablation of the *p19* locus (Fig. 1b). *p35*<sup>−/−</sup>, *p40*<sup>−/−</sup> and *p19*<sup>−/−</sup> mice, as well as relevant C57BL/6 wild-type and B6 × 129 F<sub>2</sub> *p19*<sup>+/−</sup> heterozygous controls were immunized with the encephalitogenic myelin oligodendrocyte glycoprotein 35–55 (MOG) peptide. Only *IL-23 p19*<sup>−/−</sup> and *IL-12 p40*<sup>−/−</sup> mice, the latter lacking both IL-12 and IL-23, were resistant to EAE. By contrast, *p35*<sup>−/−</sup> mice that specifically lacked IL-12 were highly susceptible to EAE (Fig. 2a). Intense mononuclear cell infiltration of the spinal cord was observed in control heterozygous *p19*<sup>+/−</sup> and *p35*<sup>−/−</sup> mice but not in *p19*<sup>−/−</sup> or *p40*<sup>−/−</sup> mice (Fig. 2b). Thus, IL-23 but not IL-12 is essential for the development of CNS autoimmune inflammation.

To determine whether exogenous IL-23 could reconstitute EAE susceptibility in the *p19*<sup>−/−</sup> and *p40*<sup>−/−</sup> mice, IL-23 gene transfer



**Figure 1** Generation of IL-23-deficient mice. See Supplementary Information for the targeting strategy. **a**, PCR analysis of wild type (+/+), heterozygous (+/−) and homozygous IL-23 p19-deficient (−/−) mice, showing presence of the targeted allele (TA) and absence of the *p19* gene in *p19*<sup>−/−</sup> mice. **b**, Quantitative RT-PCR analysis of the expression of *IL-23 p19* mRNA in lymph node tissue.



**Figure 2** IL-23, and not IL-12, is required for the induction of EAE. *p19*<sup>+/−</sup> and *p19*<sup>−/−</sup> mice were maintained on a mixed B6 × 129 F<sub>2</sub> genetic background. **a**, Average disease score of B6 wild-type (filled squares), *p19*<sup>+/−</sup> (filled circles), *p19*<sup>−/−</sup> (open squares), *p35*<sup>−/−</sup> (open diamonds) and *p40*<sup>−/−</sup> (open triangles) mice (*n* = 7–10 per group) at various days after MOG immunization. Data are representative of six experiments. **b**, CD11b immunohistochemical staining of lumbar spinal cords in mice 4 d after EAE onset. Data are representative of six experiments. Original magnification, × 200. **c**, *p19*<sup>−/−</sup> and *p40*<sup>−/−</sup> mice were treated with various rAdV gene transfer vectors or cytokines: GFP-rAdV i.c. at day 8 (open squares); IL-23-rAdV i.c. at day 8 (filled circles); IL-23-rAdV i.v. at day 8 (filled diamonds); IL-12-rAdV i.c. at day 8 (open triangles), rIL-12 i.p. from day 0 to 18 (open diamonds), rIL-12 i.p. from day 0 to 7 plus rIL-23-rAdV i.c. at day 8 (filled triangle). (Some data points are obscured owing to superimposing of the data.) **c**, Average clinical score of mice (*n* = 5 per group) treated with 3 × 10<sup>9</sup> GFP-rAdV or IL-23-rAdV either i.c. or i.v. on day 8 after MOG immunization. **d**, After MOG immunization, mice were injected as in **c** or, in addition, given 1 μg of rIL-12 i.p. daily from day 0 to day 18, 3 × 10<sup>9</sup> IL-12-rAdV i.c. at day 8, or 1 μg of rIL-12 i.p. from day 0 to day 7 plus 3 × 10<sup>9</sup> IL-23-rAdV i.c. at day 8.

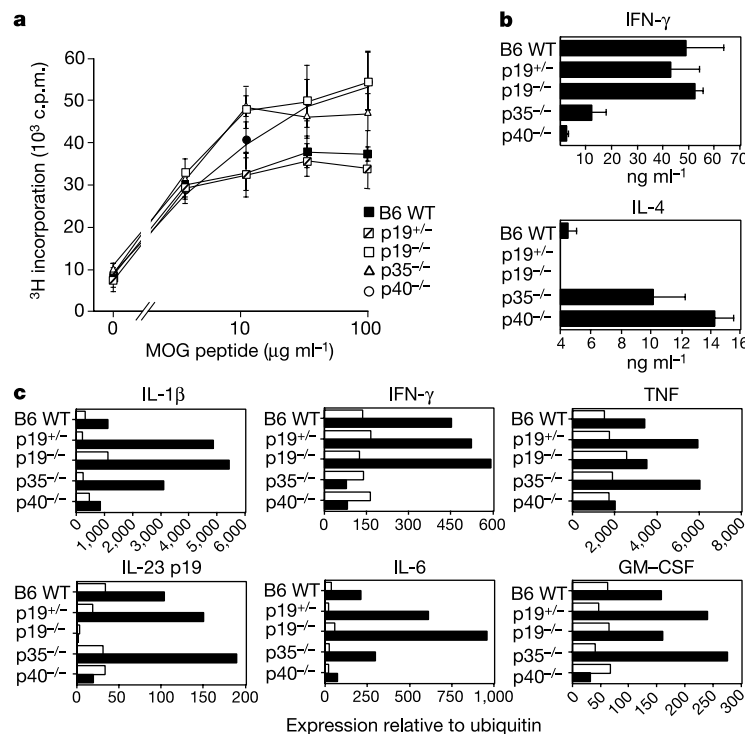
vectors were delivered into the CNS or the periphery (by intravenous (i.v.) injection) 2 d before expected disease onset. Intracerebral (i.c.) injection of gene transfer vectors has been shown to induce local CNS expression of the transgene, whereas i.v. injection induces only peripheral expression, owing to the blood-brain barrier<sup>11</sup>. IL-23 expressed in the CNS reconstituted EAE in both  $p19^{-/-}$  (Fig. 2c) and  $p40^{-/-}$  mice (Fig. 2d), although the  $p40^{-/-}$  mice had delayed disease onset and reduced disease severity. Systemic expression of IL-23 alone was not sufficient to enable disease induction (Fig. 2c, d). Injection of control adenovirus i.c. expressing only a green fluorescent protein (GFP) gene had no effect. In addition, administration of either recombinant IL-23 protein intra-peritoneally (i.p.) or IL-23 gene transfer vectors i.c. during T-cell development (from day 0 to day 6 after MOG immunization) did not reconstitute EAE in either  $p19^{-/-}$  or  $p40^{-/-}$  mice (data not shown). These data suggest that local expression of IL-23 during the effector phase of disease is required for the induction of EAE.

To assess further the contribution of IL-12 and IL-23 to the development of EAE,  $p40^{-/-}$  mice were given only IL-12 or IL-12 plus IL-23 during disease induction. Treatment of  $p40^{-/-}$  mice with recombinant IL-12 i.p. from day 0 to day 18 or IL-12 gene transfer vectors i.c. at day 8 did not facilitate disease (Fig. 2d). When  $p40^{-/-}$  mice were given IL-12 i.p. from day 0 to day 7, followed by i.c. IL-23 gene transfer at day 8, intense EAE, comparable to that seen in wild-type controls (Fig. 2a), was induced. These results suggest that IL-12 promotes development of  $T_H1$  cells, whereas IL-23 is necessary for subsequent CNS inflammatory events, for example, recruitment and/or reactivation of T cells in the CNS or activation of inflammatory and CNS-resident macrophages.

To examine these possibilities, we analysed the development of MOG-specific T cells in IL-23  $p19^{-/-}$  mice, the recruitment of

T cells to the CNS of these mice, and the effects of IL-23 on CNS-resident macrophages (microglia) and perivascular/inflammatory macrophages. Six days after MOG immunization, draining lymph node (DLN) cells from wild-type,  $p19^{+/+}$ ,  $p19^{-/-}$ ,  $p35^{-/-}$  and  $p40^{-/-}$  mice had equivalent antigen-specific proliferative responses regardless of their EAE susceptibility (Fig. 3a). DLN cells from  $p19^{-/-}$  mice secreted *in vitro* wild-type amounts of interferon- $\gamma$  (IFN- $\gamma$ ) but little or no IL-4 in response to MOG stimulation, indicating that classic  $T_H1$  cells were induced in these mice (Fig. 3b). By contrast, the response in MOG-immunized  $p35^{-/-}$  and  $p40^{-/-}$  mice was restricted to a  $T_H2$  phenotype with low IFN- $\gamma$  and high IL-4 amounts, consistent with the important role of IL-12 in the production of IFN- $\gamma$  and  $T_H1$  development<sup>12–14</sup>. The observation that  $p35^{-/-}$  mice with low IFN- $\gamma$  are susceptible to EAE, and are possibly even more severely affected than wild-type mice (Fig. 2a), is consistent with data showing that IFN- $\gamma$ -deficient mice have a hyperacute form of EAE<sup>15</sup> that seems to be related to immune dysregulation<sup>16</sup> in the presence of complete Freund's adjuvant (CFA)<sup>17</sup>.

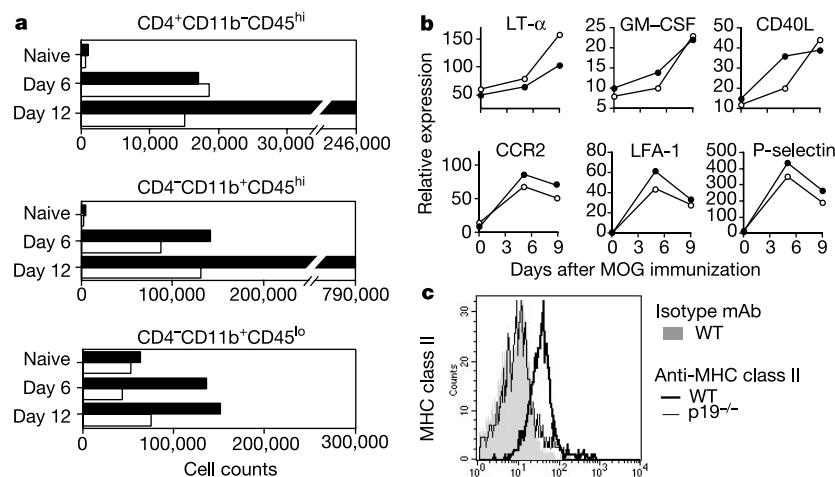
We analysed the DLNs from MOG-immunized mice by real-time quantitative PCR to determine *in vivo* expression of proinflammatory cytokines.  $p19^{-/-}$  mice produced a strong *in vivo* IFN- $\gamma$ , IL-1 $\beta$ , tumour-necrosis factor (TNF), IL-6 and granulocyte/macrophage colony-stimulating factor (GM-CSF) response, comparable to wild-type controls, whereas  $p40^{-/-}$  mice showed no induction of proinflammatory cytokines, including IL-23 p19 (Fig. 3c). DLNs from  $p35^{-/-}$  mice expressed elevated levels of IL-1 $\beta$ , TNF, IL-6, GM-CSF and IL-23 p19, but not IFN- $\gamma$ . Thus, the absence of both IL-12 and IL-23 did not prevent a T-cell proliferative response but otherwise resulted in profound immune unresponsiveness at several levels, including  $T_H1$  development and resistance to EAE induction. Absence of IL-23 alone did not prevent development or expression



**Figure 3**  $T_H1$  cells are induced in  $p19^{-/-}$  but not  $p35^{-/-}$  mice. **a**, Six days after immunization with MOG peptide emulsified in CFA, DLN cells from mice ( $n = 3$ –4 per group) were cultured in triplicate for 3 d with indicated concentrations of MOG peptide. Data are the mean  $\pm$  s.d. proliferative response of DLN cells from individual animals and are representative of three experiments. **b**, Normal expansion of IFN- $\gamma$  producing cells from  $p19^{-/-}$  mice. DLN cells (6 d after MOG immunization) from individual mice were

cultured separately for 60 h with MOG peptide, and the supernatants were tested for secreted cytokines by enzyme-linked immunosorbent assay. Data are the mean  $\pm$  s.d. response of 3–4 mice and are representative of two experiments. **c**, Cytokine mRNA expression of DLN from mice immunized with MOG. DLN cells from naive mice (open bars) or mice 6 d after immunization (filled bars) were prepared for real-time PCR analysis. Data shown are pooled samples of 5–8 mice from two experiments.





**Figure 4** T<sub>H</sub>1 cells and inflammatory macrophages enter the CNS of MOG-immunized *p19*<sup>-/-</sup> mice but fail to induce EAE. **a**, Number of CD4<sup>+</sup>CD45<sup>hi</sup> T cells, CD11b<sup>+</sup>CD45<sup>hi</sup> inflammatory macrophages and CD11b<sup>+</sup>CD45<sup>lo</sup> CNS-resident microglia recovered from the brain and spinal cord of naive and MOG-immunized wild-type (filled bars) and *p19*<sup>-/-</sup> (open bars) mice. **b**, Expression of proinflammatory cytokine mRNA in spinal cord of MOG immunized wild-type (filled circles) and *p19*<sup>-/-</sup> (open circles) mice before onset

of EAE. Cytokine mRNA levels were determined by real-time PCR analysis of pooled, whole spinal cord samples from 3–5 mice per group. Results are representative of two experiments. **c**, Resident microglia expression of MHC class II molecules. Flow cytometric analysis of MHC class II molecules was done on gated CD4<sup>-</sup>CD11b<sup>+</sup>CD45<sup>lo</sup> resident microglia. Results are representative of three experiments.

of T<sub>H</sub>1-associated proinflammatory cytokines, but did prevent EAE. These results indicate that IL-23 is required for steps of disease development subsequent to initial T-cell activation.

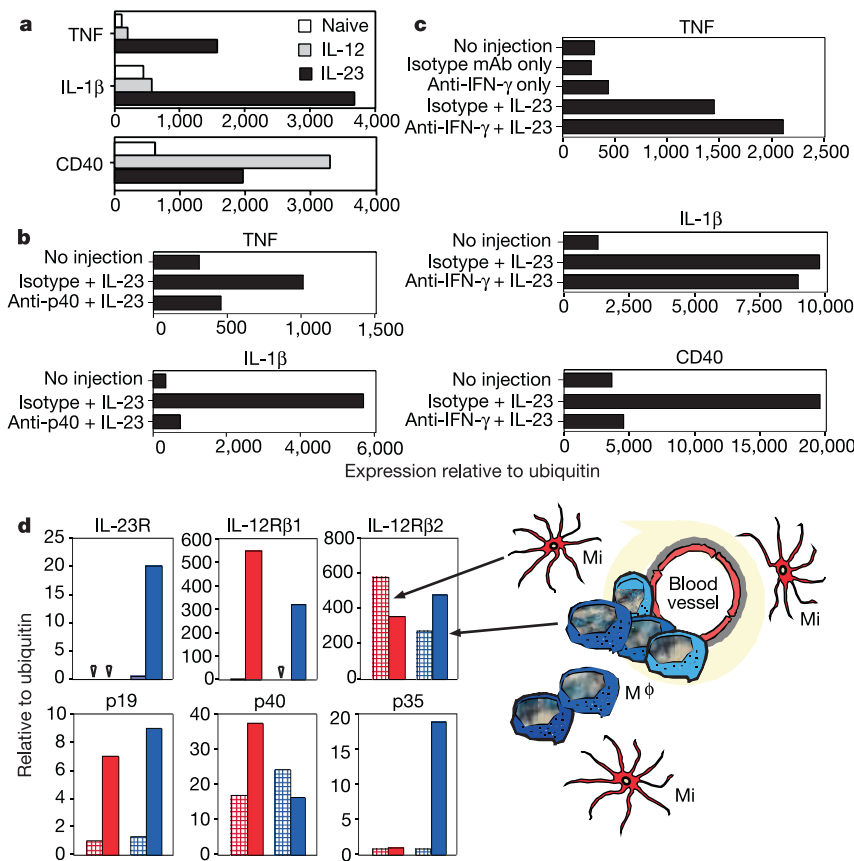
We next determined whether activated T cells from *p19*<sup>-/-</sup> mice could infiltrate the CNS. During EAE pathogenesis, CD4<sup>+</sup> T cells and CD11b<sup>+</sup> monocytes enter the CNS well before the onset of clinical disease<sup>18,19</sup>. These infiltrating cells are characterized by high CD45 expression, whereas resident brain macrophages (or microglia) are CD11b<sup>+</sup>CD45<sup>lo</sup> (ref. 19 and see Supplementary Fig. 2). Six days after MOG immunization, increased and comparable numbers of CD4<sup>+</sup> T cells and CD11b<sup>+</sup> macrophages were recovered from the whole CNS of both *p19*<sup>-/-</sup> and wild-type control mice (Fig. 4a). Quantitative real-time PCR analysis of spinal cord mRNA showed that 9 d after immunization there was comparable expression of lymphotoxin-α, GM-CSF, CD40L, lymphocyte function associated molecule-1, P-selectin and CCR2 mRNA (all of which are expressed predominantly by T cells during invasion into the CNS) in both the *p19*<sup>-/-</sup> and wild-type mice (Fig. 4b). Twelve days after immunization, inflamed spinal cords from wild-type mice with EAE showed a 200- to 250-fold increase in CD4<sup>+</sup> cells and CD4<sup>-</sup>CD11b<sup>+</sup>CD45<sup>hi</sup> macrophages, whereas the EAE-resistant *p19*<sup>-/-</sup> mice had considerably fewer infiltrating T cells and macrophages (Fig. 4a). In addition, CD4<sup>-</sup>CD11b<sup>+</sup>CD45<sup>lo</sup> microglia numbers did not increase in *p19*<sup>-/-</sup> mice after immunization (Fig. 4a) and failed to upregulate major histocompatibility complex (MHC) class II molecules (Fig. 4c). Thus, in the absence of IL-23, T<sub>H</sub>1 cells entered the CNS, but their presence did not lead to further recruitment of T cells or macrophages, or activation of resident microglia.

The endotoxic shock-like effects of transgenic IL-23 in mice<sup>10</sup>, the expression of IL-23R complex (IL-12Rβ1 and IL-23R) in bone-marrow-derived macrophages<sup>9</sup>, and the finding that subsets of macrophages and other cell types express Stat4 but do not respond to IL-12 (ref. 20) together suggest that IL-23 might have direct effects on myeloid cells. To test this in a simple model system, we administered IL-23, using IL-12 as a control cytokine, into the peritoneum of mice and analysed peritoneal macrophage gene expression by quantitative real-time PCR. Three hours after cytokine injection, IL-23 but not IL-12 treatment induced macrophages (F4/80<sup>+</sup>CD11b<sup>+</sup>CD11c<sup>-</sup>B220<sup>-</sup>) to express *IL-1β* and *TNF* mRNA. By contrast, both IL-12 and IL-23 increased the expression of *CD40* (Fig. 5a) and several other inflammatory molecules, including

*MMP2*, *MMP7*, *MMP9* and *iNOS* (data not shown). Co-injection of a monoclonal antibody against p40 together with the cytokine, IL-23, completely blocked the induction of *IL-1β* and *TNF* mRNA (Fig. 5b), indicating the specific induction of these proinflammatory cytokines by IL-23. Conversely, a 1-h pretreatment of mice with monoclonal antibodies against IFN-γ before IL-23 injection blocked the increase in expression of *CD40* but not *IL-1β* or *TNF* (Fig. 5c). IL-23 also induced expression of *IL-1β* and *TNF* but not *CD40* in peritoneal macrophages in IFN-γ-deficient mice (data not shown). These results suggest that IL-23 directly activates macrophages *in vivo* to express IL-1β and *TNF*, but that expression of *CD40* and other proinflammatory mediators such as MMPs and *iNOS* is IFN-γ dependent. The source of IL-23-induced IFN-γ has not been defined.

To analyse the potential of myeloid cells in the CNS to produce and to respond to IL-23 and IL-12, microglia and inflammatory macrophages entering the CNS were sorted to purity and assessed for expression of IL-23 and IL-12 receptor and ligand genes using real-time quantitative-PCR approaches. Cells were isolated 2 d before clinical onset of EAE or 2 d after clinical onset when mice were severely affected. During EAE, *IL-23R* and *IL-12Rβ1*, which form the functional IL-23 signalling complex, were coexpressed by inflammatory macrophages. By contrast, resident microglia expressed only *IL-12Rβ1* and not *IL-23R* (Fig. 5d). No detectable levels of *IL-23R* or *IL-12Rβ1* mRNA were found in resident microglia or inflammatory macrophages 2 d before EAE onset. *IL-12Rβ2* mRNA was expressed constitutively by resident and inflammatory macrophages both before and after EAE onset (Fig. 5d), as well as by cells isolated from the naive CNS (data not shown). IL-23 ligand subunits (*p19* and *p40*) were coexpressed on both the resident microglia and inflammatory macrophages in CNS 2 d before disease onset. After disease onset, the *p19* mRNA levels increased markedly on both the microglia and macrophage populations (Fig. 5d). IL-12 ligand subunits (*p35* and *p40*) were coexpressed only in inflammatory macrophages after disease onset. These results suggest that IL-23 is produced by both resident microglia and inflammatory macrophages, but that only inflammatory macrophages can respond to IL-23. In contrast, IL-12 is produced primarily by inflammatory macrophages, but both microglia and macrophages can potentially respond to IL-12.

The current study clearly shows that the *in vivo* functions



**Figure 5** IL-23, and not IL-12, induces *IL-1β* and *TNF* mRNA expression in macrophages. **a**, C57BL/6 wild-type mice ( $n = 5$  per group) were given 5  $\mu$ g of IL-12 or IL-23 i.p. Three hours after treatment, F4/80<sup>+</sup>CD11b<sup>+</sup>CD11c<sup>+</sup>B220<sup>+</sup> peritoneal macrophages were sorted and prepared for quantitative real-time PCR analysis. Results are representative of three experiments. **b**, Mice ( $n = 3$  per group) were injected with a 10-molar excess of monoclonal antibodies against p40 plus 5  $\mu$ g of IL-23. Data are representative of two experiments. **c**, Mice ( $n = 3$ –4 per group) were given by i.p. injection 1 mg of anti-IFN- $\gamma$  or isotype control monoclonal antibodies 1 h before i.p. administration of 5  $\mu$ g of IL-23.

Three hours after IL-23 treatment, macrophages were isolated, sorted to purity and prepared for quantitative real-time PCR analysis. Results are representative of two experiments. **d**, Left, microglia (red bars) and CNS inflammatory macrophages (blue bars) 2 d before disease onset (checked bars) and 2 days after disease onset (filled bars), were isolated and tested for gene expression by quantitative PCR. Primer sets were specific for the indicated genes. Right, cartoon showing the location of microglia (Mi) in the CNS parenchyma and the inflammatory macrophages (M $\phi$ ) invading the CNS across the perivascular spaces.

previously attributed to IL-12 for EAE are induced by two distinct cytokines: IL-12 promotes the development of naive T cells, and IL-23 mediates late-stage inflammation. However, while IL-12 may be sufficient for the growth and development of T<sub>H</sub>1 cells, other factors such as IFN- $\alpha$  (ref. 21), osteopontin<sup>22</sup> and IL-27 (refs 23, 24) may partially replace IL-12 by inducing IFN- $\gamma$  production and T<sub>H</sub>1 responses in certain microbial infections, allograft rejection and autoimmunity<sup>25–27</sup>. By contrast, IL-23 seems to be necessary for maintaining chronic inflammation (Figs 2, 4). The expression of IL-23R on CNS inflammatory macrophages and the ability of IL-23 to activate macrophages to produce proinflammatory cytokines such as IL-1 and TNF suggest that it is an important regulator of myeloid cells in organ-specific autoimmune diseases. In addition, IL-23 has potent activities on memory-activated T cells<sup>8</sup>, suggesting that IL-23 is a critical and broad regulator of late-stage inflammatory processes. Thus, therapeutic targeting of IL-23 or IL-23-receptor may be a promising approach for a range of inflammatory autoimmune diseases. □

## Methods

### Mice and immunization

IL-12 p35- and p40-deficient mice obtained from Jackson Laboratory were originally generated on the B6  $\times$  129 background and were backcrossed 11 generations onto the C57BL/6 background. IL-23 p19-deficient mice were generated using approaches as

described<sup>28</sup>, and *p19*<sup>−/−</sup> *p19*<sup>+/−</sup> and *p19*<sup>+/+</sup> wild-type controls were maintained on a mixed B6  $\times$  129 F<sub>2</sub> background. The induction of EAE with MOG35–55 peptide in CFA plus pertussis toxin, and the immunohistochemical assessment of fresh spinal cord tissues are described in the Supplementary Information. Draining lymph node (DLN) cells were isolated 6 d after immunization. The T-cell proliferation and cytokine secretion assays are described in the Supplementary Information.

### IL-23 and IL-23-rAdV treatment

We produced IL-23 and IL-23-rAdV as described<sup>8</sup>. For administration into the CNS, recombinant cytokines or rAdV gene transfer vectors were suspended in 10  $\mu$ l of PBS and injected into the right-lateral cerebral ventricle using a 28-gauge needle as described<sup>11</sup>. For peripheral delivery, cytokines or rAdV vectors suspended in 100  $\mu$ l of PBS were injected into the tail vein.

### Isolation of CNS mononuclear cells

T cells, macrophages and resident microglia were isolated by digesting brain and/or spinal cord homogenate with collagenase and DNase followed by Percoll gradient centrifugation as described<sup>19</sup>. We determined the number of CD4<sup>+</sup>CD45<sup>hi</sup> T cells, CD4<sup>−</sup>CD11b<sup>+</sup>CD45<sup>hi</sup> inflammatory macrophages and CD4<sup>−</sup>CD11b<sup>+</sup>CD45<sup>lo</sup> resident microglia in the CNS by multiplying the percentage of lineage-marker-positive cells by the total number of mononuclear cells isolated from the CNS (see Supplementary Fig. 2 for typical flow cytometric analysis of these preparations). Inflammatory macrophages and resident microglia were isolated from irradiation bone marrow chimaeric mice produced by the injection of CD45.1 B6 bone marrow cells into irradiated CD45.2 B6 mice. The CD11b<sup>+</sup>CD45.1<sup>hi</sup>CD45.2<sup>−</sup> inflammatory macrophages and CD11b<sup>+</sup>CD45.2<sup>lo</sup>CD45.1<sup>−</sup> resident microglia (radiation-resistant cells remaining with the host CD45 allotype) were purified by three-colour flow cytometric sorting from more than 100 irradiation bone-marrow chimaeric mice<sup>19,29</sup>. At various time points after EAE induction, purified and sorted cells were pooled (microglia or inflammatory macrophages) and analysed by

quantitative real-time PCR for the expression of IL-12 and IL-23 ligand and receptor subunits.

## Gene expression analysis

We extracted RNA from tissues or cell pellets using RNeasy columns (Qiagen) and treated them with Dnase I (Promega). Complementary DNAs were prepared as described<sup>11</sup> and used as templates for quantitative PCR. We analysed 25 ng of cDNA for expression of a range of genes using the GeneAmp 5700 Sequence Detection System (Applied Biosystems). Analysis of cDNA samples from spinal cords, draining lymph nodes or peritoneal macrophages was normalized to expression of the housekeeping gene, ubiquitin.

## In vivo macrophage stimulation

Mice received an i.p. injection of IL-12 or IL-23 in 100 µl of PBS. Where indicated, mice were pretreated with 1 mg of neutralizing rat monoclonal antibodies against mouse IFN-γ (XMG1.2) or against β-galactose (β-gal; isotype control, rat IgG1) 1 h before cytokine treatment. For IL-23 blocking studies, mice were injected i.p. with a 10-molar excess of neutralizing rat antibodies against mouse p40 (C17.8) or against β-gal (isotype control, rat IgG2a) at the time of IL-23 administration. Three hours after cytokine treatment, peritoneal macrophages (F4/80<sup>+</sup> (monoclonal antibody F4/80), CD11b<sup>+</sup> (monoclonal antibody M1/70), CD11c<sup>+</sup> (monoclonal antibody HL3) and CD45R/B220<sup>+</sup> (monoclonal antibody RA3-6B2, BD Biosciences) were isolated by multicolour flow cytometry and prepared for real-time quantitative PCR analysis.

Received 10 September; accepted 14 November 2002; doi:10.1038/nature01355.

- Kobayashi, M. *et al.* Identification and purification of natural killer cell stimulatory factor (NKSF), a cytokine with multiple biologic effects on human lymphocytes. *J. Exp. Med.* **170**, 827–845 (1989).
- Trinchieri, G. *et al.* Natural killer cell stimulatory factor (NKSF) or interleukin-12 is a key regulator of immune response and inflammation. *Prog. Growth Factor Res.* **4**, 355–368 (1992).
- Leonard, J. P., Waldburger, K. E. & Goldman, S. J. Prevention of experimental autoimmune encephalomyelitis by antibodies against interleukin 12. *J. Exp. Med.* **181**, 381–386 (1995).
- Issazadeh, S., Ljungdahl, A., Hojeberg, B., Mustafa, M. & Olsson, T. Cytokine production in the central nervous system of Lewis rats with experimental autoimmune encephalomyelitis: dynamics of mRNA expression for interleukin-10, interleukin-12, cytolysin, tumour necrosis factor α and tumour necrosis factor β. *J. Neuroimmunol.* **61**, 205–212 (1995).
- Leonard, J. P., Waldburger, K. E. & Goldman, S. J. Regulation of experimental autoimmune encephalomyelitis by interleukin-12. *Ann. NY Acad. Sci.* **795**, 216–226 (1996).
- Constantinescu, C. S. *et al.* Antibodies against IL-12 prevent superantigen-induced and spontaneous relapses of experimental autoimmune encephalomyelitis. *J. Immunol.* **161**, 5097–5104 (1998).
- Segal, B. M., Dwyer, B. K. & Shevach, E. M. An interleukin (IL)-10/IL-12 immunoregulatory circuit controls susceptibility to autoimmune disease. *J. Exp. Med.* **187**, 537–546 (1998).
- Oppmann, B. *et al.* Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. *Immunity* **13**, 715–725 (2000).
- Parham, C. *et al.* A receptor for the heterodimeric cytokine IL-23 is composed of IL-12Rβ1 and a novel cytokine receptor subunit, IL-23R. *J. Immunol.* **168**, 5699–5708 (2002).
- Wiekowski, M. T. *et al.* Ubiquitous transgenic expression of the IL-23 subunit p19 induces multiorgan inflammation, runting, infertility, and premature death. *J. Immunol.* **166**, 7563–7570 (2001).
- Cua, D. J., Hutchins, B., LaFace, D. M., Stohlman, S. A. & Coffman, R. L. Central nervous system expression of IL-10 inhibits autoimmune encephalomyelitis. *J. Immunol.* **166**, 602–608 (2001).
- O'Garra, A. & Arai, N. The molecular basis of T helper 1 and T helper 2 cell differentiation. *Trends Cell Biol.* **10**, 542–550 (2000).
- Caspi, R. IL-12 in autoimmunity. *Clin. Immunol. Immunopathol.* **88**, 4–13 (1998).
- Falcone, M. & Sarvetnick, N. Cytokines that regulate autoimmune responses. *Curr. Opin. Immunol.* **11**, 670–676 (1999).
- Willenborg, D. O., Fordham, S., Bernard, C. C., Cowden, W. B. & Ramshaw, I. A. IFN-γ plays a critical down-regulatory role in the induction and effector phase of myelin oligodendrocyte glycoprotein-induced autoimmune encephalomyelitis. *J. Immunol.* **157**, 3223–3227 (1996).
- Chu, C. Q., Wittmer, S. & Dalton, D. K. Failure to suppress the expansion of the activated CD4 T cell population in interferon-γ-deficient mice leads to exacerbation of experimental autoimmune encephalomyelitis. *J. Exp. Med.* **192**, 123–128 (2000).
- Matthys, P., Vermeire, K. & Billiau, A. Mac-1<sup>+</sup> myelopoiesis induced by CFA: a clue to the paradoxical effects of IFN-γ in autoimmune disease models. *Trends Immunol.* **22**, 367–371 (2001).
- Hickey, W. F., Hsu, B. L. & Kimura, H. T-lymphocyte entry into the central nervous system. *J. Neurosci. Res.* **28**, 254–260 (1991).
- Sedgwick, J. D. *et al.* Isolation and direct characterization of resident microglial cells from the normal and inflamed central nervous system. *Proc. Natl Acad. Sci. USA* **88**, 7438–7442 (1991).
- Frucht, D. M. IL-23: a cytokine that acts on memory T cells. *Sci. STKE* **114** (2002) pe1 [online] (<http://stke.sciencemag.org/cgi/content/full/sigtrans>) (2002).
- Nguyen, K. B. *et al.* Critical role for STAT4 activation by type 1 interferons in the interferon-γ response to viral infection. *Science* **297**, 2063–2066 (2002).
- Chabas, D. *et al.* The influence of the proinflammatory cytokine, osteopontin, on autoimmune demyelinating disease. *Science* **294**, 1731–1735 (2001).
- Chen, Q. *et al.* Development of Th1-type immune responses requires the type I cytokine receptor TCCR. *Nature* **407**, 916–920 (2000).
- Pflanz, S. *et al.* IL-27, a heterodimeric cytokine composed of EB13 and p28 protein, induces proliferation of naive CD4<sup>+</sup> T cells. *Immunity* **16**, 779–790 (2002).
- Holscher, C. *et al.* A protective and agonistic function of IL-12p40 in mycobacterial infection. *J. Immunol.* **167**, 6957–6966 (2001).
- Piccotti, J. R. *et al.* Alloantigen-reactive Th1 development in IL-12-deficient mice. *J. Immunol.* **160**, 1132–1138 (1998).
- Becher, B., Durell, B. G. & Noelle, R. J. Experimental autoimmune encephalitis and inflammation in the absence of interleukin-12. *J. Clin. Invest.* **110**, 493–497 (2002).
- Chen, S. C. *et al.* Impaired pulmonary host defense in mice lacking expression of the CXCL chemokine lungkine. *J. Immunol.* **166**, 3362–3368 (2001).

29. Hickey, W. F. Migration of hematogenous cells through the blood-brain barrier and the initiation of CNS inflammation. *Brain Pathol.* **1**, 97–105 (1991).

**Supplementary Information** accompanies the paper on Nature's website (<http://www.nature.com/nature>).

**Acknowledgements** We thank J. Cupp and the DNAX FACS Facility; and K. Moore, G. Zurawski and D. Rennick for comments. DNAX Research Inc. is supported by Schering-Plough.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to D.J.C. (e-mail: [daniel.cua@dnax.org](mailto:daniel.cua@dnax.org)).

## Targeted recycling of PECAM from endothelial surface-connected compartments during diapedesis

Zahra Mamdouh\*, Xia Chen\*, Lynda M. Pierini†, Frederick R. Maxfield† & William A. Muller\*

\* Departments of Pathology and Laboratory Medicine, and † Biochemistry, Weill Medical College, 1300 York Avenue, New York, New York 10021, USA

Leukocytes enter sites of inflammation by squeezing through the borders between endothelial cells that line postcapillary venules at that site. This rapid process, called transendothelial migration (TEM) or diapedesis, is completed within 90 s after a leukocyte arrests on the endothelial surface<sup>1–4</sup>. In this time, the leukocyte moves in amoeboid fashion across the endothelial borders, which remain tightly apposed to it during transit. It is not known how the endothelial cell changes its borders rapidly and reversibly to accommodate the migrating leukocyte. Here we show that there is a membrane network just below the plasmalemma at the cell borders that is connected at intervals to the junctional surface. PECAM-1, an integral membrane protein with an essential role in TEM<sup>5–7</sup>, is found in this compartment and constitutively recycles evenly along endothelial cell borders. During TEM, however, recycling PECAM is targeted to segments of the junction across which monocytes are in the act of migration. In addition, blockade of TEM with antibodies against PECAM specifically blocks the recruitment of this membrane to the zones of leukocyte migration, without affecting the constitutive membrane trafficking.

Intact human umbilical vein endothelial cell (HUVEC) monolayers stained with fluorescein isothiocyanate (FITC)-conjugated monoclonal antibodies against PECAM showed a thin line of fluorescence along the cell borders with accentuation at points where cells overlap (Fig. 1a). The same procedure carried out on permeabilized monolayers (Fig. 1b) showed a broad band of PECAM-bearing reticular structures beneath each border. We refer to these structures descriptively as the 'subjunctional reticulum'. PECAM-bearing membrane from the cell border was rapidly internalized into this reticulum (Fig. 1c). VE-cadherin and CD99, two other membrane proteins that localize to the intercellular border<sup>8</sup>, showed a similar pattern (Fig. 1d, e), whereas diffusely distributed proteins such as the adhesion molecule ICAM-1 (Fig. 1f) remained on the endothelial apical surface for at least 1 h, and at later time points were detected in perinuclear vesicles. In addition, soluble endocytic tracers such as FITC-labelled dextran did not detectably enter these subjunctional compartments within this time period (data not shown).



The subjunctional reticulum of PECAM-bearing membrane did not resemble typical recycling endosomes and did not colocalize with typical endocytic markers such as the transferrin receptor (Supplementary Fig. 1S). To determine whether PECAM-bearing membrane was internalized into true vesicles sealed off from the cell surface, we made use of the fact that fluorescence of FITC is pH sensitive. We incubated HUVEC monolayers with FITC-labelled Fab fragments of monoclonal antibodies against PECAM at 37 °C to saturate both the sequestered and surface PECAM compartments, and then lowered the temperature to 4 °C. Control monolayers were labelled at 4 °C to restrict labelling to the cell surface. Digitized images were taken before and immediately after changing to a buffer system that has been shown to lower extracellular pH without affecting the pH of internal cellular compartments<sup>9</sup>. The surface fluorescence was quenched as expected, but unexpectedly the fluorescence in the sequestered compartment was also quenched to the same extent (Fig. 2a, b). Thus, although molecules as small as Fab fragments of IgG were excluded from this compartment at 4 °C, protons could enter, indicating that these were not sealed vesicles.

At the ultrastructural level, PECAM was detected diffusely along the intercellular borders whether incubation was carried out at 37 °C or on ice (Fig. 2c–g), confirming our observations by light microscopy (Fig. 1) and showing that the antibody had full access to the junctions. In cells incubated at 37 °C, we observed numerous 3–3'-diaminobenzidine tetrahydrochloride (DAB)-positive vesicle-like structures (about 50 nm in diameter); many of these showed visible connections to each other and were in continuity with the cell border as well as located many diameters away from the border (Fig. 2c, d). Most of these appeared to be completely pinched off from the cell surface; only the FITC quenching experiment showed that they formed a system of surface-connected membrane compartments. At 4 °C, abundant vesicle-like structures with the same morphology were present, but almost all were devoid of DAB reaction product, showing that incubation with the antibody did not induce this compartment and that the antibody did not have access when cells were incubated on ice (Fig. 2f, g).

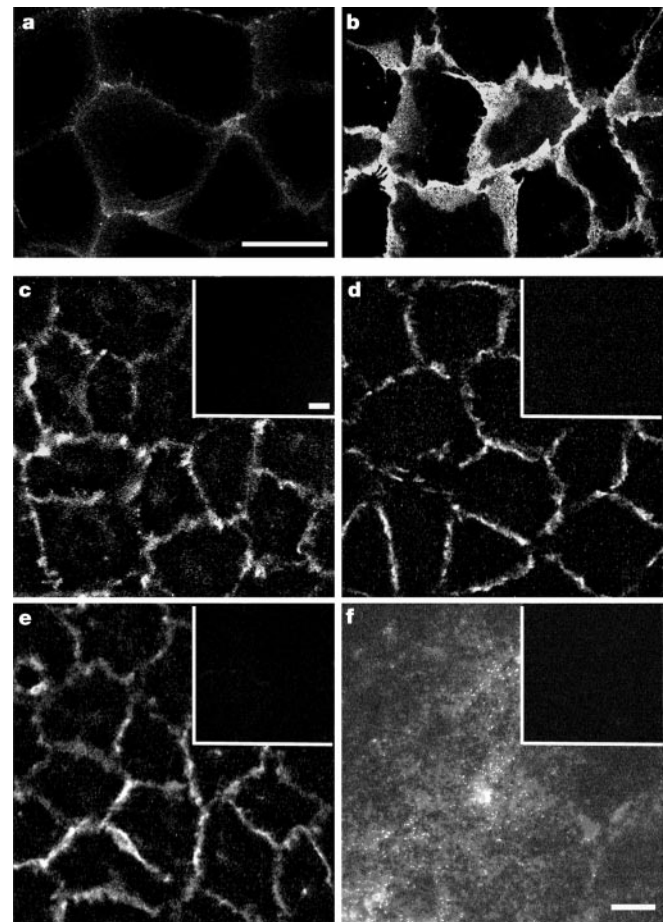
Stereological analysis of electron micrographs indicated that the ratio of labelled membrane in the subjunctional reticulum to labelled membrane at the cell border was 0.44. Consistent with this, comparison of digitized images of cells labelled at 4 °C and 37 °C showed that fluorescence of the latter was about 1.5-fold higher. Therefore, roughly a third of the total cell PECAM was in this sequestered pool at any time.

To determine whether the PECAM-bearing membrane is recycled constitutively from these surface-connected compartments to the cell border, we designed a procedure to label a recycling pool of PECAM (see Methods). At time zero, there was no staining of the intercellular border, indicating that all of the internalized PECAM was still inaccessible to secondary antibody. Within 5 min of warming to 37 °C, however, fluorescence began to appear at the borders between adjacent cells, showing the return of previously internalized PECAM-bearing membrane to the endothelial junctions. Fluorescence intensity increased with time and reached a plateau by 30 min (Fig. 3a, b). Independent experiments in which externalization of PECAM was analysed by flow cytometry showed a similar plateau within 15–30 min (Fig. 3c).

We repeated the PECAM recycling studies in the presence of transmigrating monocytes, arresting the process by paraformaldehyde fixation to catch monocytes in the act of diapedesis. PECAM-bearing membrane recycled evenly along cell borders in locations that were free from monocytes, as judged by the uniform fluorescence intensity along the junctions. Where monocytes were present, however, rather than recycling evenly along the cell border, recycling PECAM now preferentially surrounded monocytes that were actively transmigration (Fig. 4a). Note that in these experiments, we labelled only recycling endothelial cell PECAM, not total PECAM, which remained unchanged. This selective recycling of

PECAM-bearing membrane to the zone of transmigration was seen for monocytes at all stages of diapedesis: those just entering the junction, those midway through, and those almost completely across the endothelial junction (Fig. 4a). We favour the idea that recycling membrane is directly targeted to the region of the cell border where diapedesis is occurring. In support of this, experiments visualizing total PECAM in junctions during TEM showed a loss of network staining and a narrow band of intense fluorescence adjacent to the monocyte (Supplementary Table 1 and Fig. 5S). The selective enrichment was not due to the 'trapping' of recycled PECAM by homophilic interactions with PECAM on the monocytes, as we repeated the same experiments with the THP-1 monocyte line, which express even more PECAM than monocytes, adhere to the endothelial cells and move to the junctions, but cannot transmigrate. In the absence of transmigration, the THP-1 cells remained bound at the HUVEC borders, but PECAM recycled evenly along the junction without enrichment near the PECAM-bearing THP-1 cells (Fig. 4b).

The preceding experiments using PECAM only as a membrane

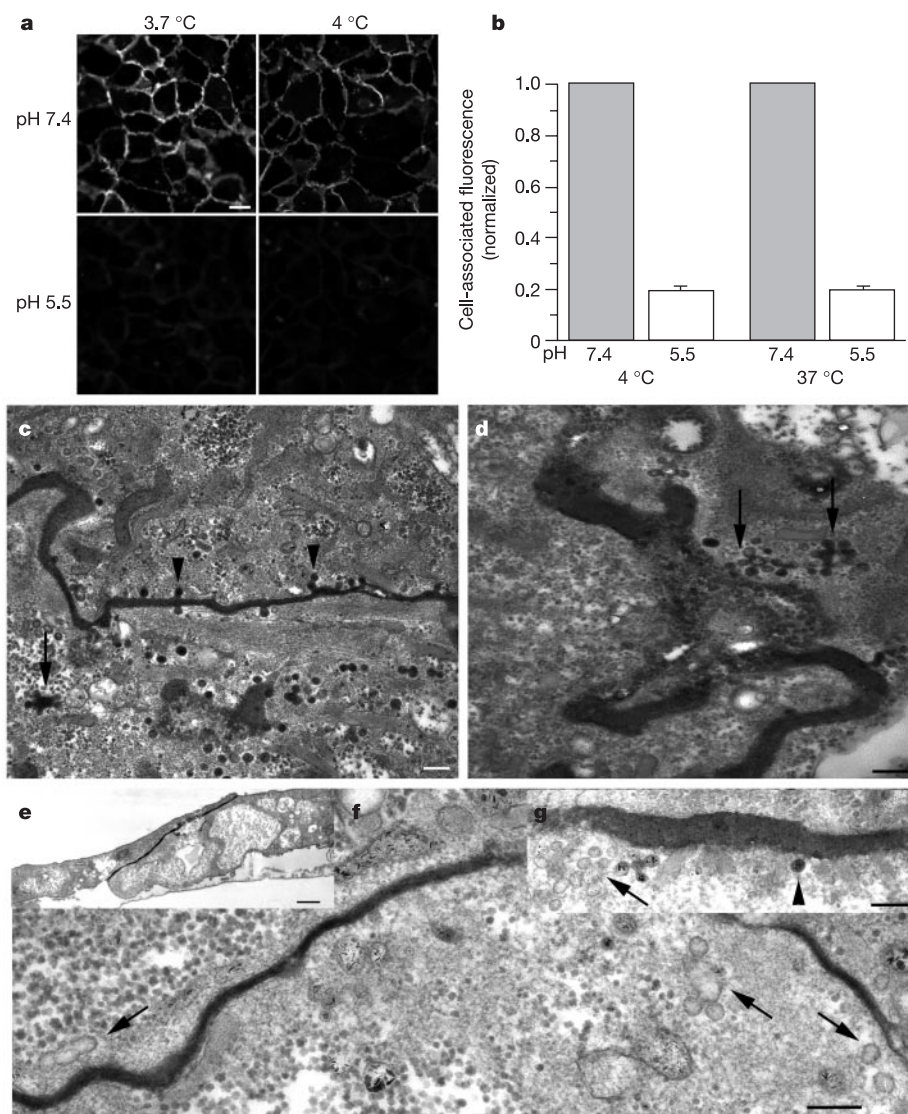


**Figure 1** PECAM-rich membrane is rapidly internalized from the cell border into a subjunctional reticulum. Confluent HUVEC monolayers were fixed in paraformaldehyde (a) or fixed and permeabilized (b), immunostained with FITC-labelled antibodies against PECAM and examined by confocal microscopy. c–f, Confluent HUVEC monolayers were incubated for 1 h at 37 °C with monoclonal antibodies against PECAM-1 (c), VE-cadherin (d), CD99 (e) or ICAM-1 (f), and then washed and fixed in paraformaldehyde. After blocking externally disposed monoclonal antibodies with unlabelled RAM (100 μg ml<sup>-1</sup>, 1 h, 4 °C), the cells were permeabilized, and internalized membrane was visualized using FITC–RAM. Replicate monolayers that were not permeabilized but exposed to FITC–RAM (insets) show that this procedure successfully blocks the detection of antigens remaining on the cell surface. Scale bar, 20 μm.

marker nonetheless suggested that a productive interaction of leukocyte PECAM with endothelial cell PECAM might be involved in targeted membrane delivery during TEM. We repeated the above experiments using monocytes pretreated with a blocking antibody against PECAM or preimmune IgG. The preimmune IgG had no effect on the transmigration or recycling of internalized PECAM (Fig. 4a). By contrast, when monocyte PECAM was blocked, monocytes adhered to the endothelial monolayer and moved to the junctions, but could not transmigrate, as described previously<sup>5,7,10</sup>. When monocyte PECAM was blocked, endothelial cell PECAM recycled evenly, without enrichment in the zone around the monocytes at the apical surface of the monolayer (Fig. 4c). Our results suggest that homophilic interaction between

leukocyte PECAM and endothelial cell PECAM is necessary (Fig. 4c) but not sufficient (Fig. 4b) to trigger or sustain targeted PECAM recycling.

These results explain the previously published observations that preincubation of HUVEC cultures on ice with monoclonal antibodies against PECAM does not block TEM<sup>5</sup>, whereas preincubation with the same antibody at 37 °C does<sup>8</sup>. On ice, the monoclonal antibody had access to the surface pool of PECAM at the cell border, but saturating this pool had no effect on TEM. Only when the monoclonal antibody had access to the sequestered compartment at 37 °C was it effective at blocking TEM. This shows that PECAM in the surface-connected compartments of endothelial cells is essential for transmigration.



**Figure 2** PECAM-bearing membrane enters a reticulum of surface-connected compartments. **a**, HUVEC monolayers were incubated with FITC-labelled Fab fragments against PECAM at 37 °C or 4 °C, washed and chilled, and then immunofluorescence images were obtained from the monolayers at pH 7.4. The extracellular pH was lowered to 5.5 using a cell-impermeable buffer, and the image of the same field was immediately obtained. **b**, Quantification of a series of digital images as in **a**. **c–g**, Immunoelectron microscopy of HUVEC incubated with horseradish peroxidase (HRP)-labelled monoclonal antibodies against PECAM. **c, d**, In cells incubated at 37 °C, DAB reaction product shows that antibodies specific for PECAM enter an extensive series of vesicle-like structures that

interconnect with each other (arrows), as well as with the cell surface at the intercellular junction (arrowheads). Cells incubated at 4 °C (**e–g**) show DAB reaction products along the intercellular border. Abundant interconnected vesicle-like structures are seen adjacent to the junctions (arrows); however, only rare profiles bear DAB reaction product (arrowhead). No structures were visualized when cells were incubated with free HRP at the same concentration at either temperature. Scale bars: 20 μm (**a**); 200 nm (**c, d, f, g**); and 500 nm (**e**). Images in **c, d, f** and **g** are *en face* sections parallel to the substratum; image in **e** is a cross-section with the apical surface facing up.

We have identified a new membrane compartment at the intercellular borders of endothelial cells. PECAM-bearing membrane constitutively fluxes into and out of an extensive network of vesicle-like structures located just beneath the plasma membrane. These vesicles are connected to each other and to the cell surface. Although these membranes resemble caveolae superficially, caveolae are rare in cultured endothelial cells and, when present, are seen mostly at the apical and basal surfaces, and not concentrated at the cell borders. PECAM and caveolin 1 did not colocalize in our HUVEC cultures nor did they co-migrate on sucrose gradients (Supplementary Figs 2S–4S).

Likewise, we do not think that this compartment represents the vesiculo-vacuolar organelles that have been reported<sup>11</sup> to mediate permeability increases in endothelium in response to vascular endothelial growth factor and other factors. Whereas vesiculo-vacuolar organelles tend to accumulate near endothelial cell borders, they are not present in cultured HUVEC<sup>12</sup>, they contain caveolin and they fuse with the apical and basal surfaces of endothelium rather than showing the junctional fusion that we describe. A compartment of surface-connected membranes, albeit of different morphology and function, has been described in PC12 cells<sup>13</sup>. This “sub-plasmalemmal tubulo-cisternal membrane system” forms a membrane compartment that is distinct from caveolae and endosomes, and is accessible to a membrane-impermeable reducing agent, MesNa, at 4 °C, but not to the protein avidin<sup>13</sup>.

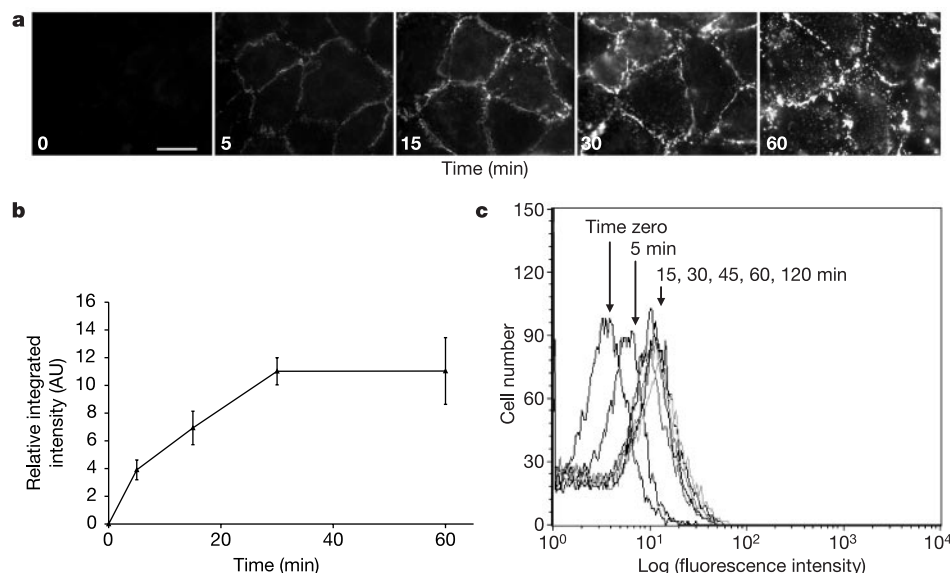
The most significant finding of this study is that TEM involves the recycling of PECAM targeted to the region of the junction where migration is taking place. Although we have measured only movement of PECAM itself, the most reasonable mechanism would be one that involves movement of a PECAM-bearing membrane. Whether other membrane proteins and lipids move with PECAM is currently under investigation.

Although it is clear that most TEM occurs at cell junctions<sup>4</sup>, there are reports of trafficking of leukocytes through endothelium, close to but not at the endothelial borders<sup>14–16</sup>. Although our data do not address directly the mode of transmigration, as a unifying hypothesis, we might speculate that under certain

conditions the PECAM-containing membranes of the subjunctional reticulum could provide the lining for an alternative or parallel junction lined by PECAM and the other junction-specific molecules that comprise this membrane compartment. Considering that there is almost half as much membrane in this compartment as at the cell border, this would be physically possible.

Monocyte TEM is associated with a transient disruption of the VE-cadherin–catenin complex at the site of and coincident with monocyte migration<sup>17,18</sup>. Although this seems to be at odds with the enrichment of recycling PECAM in this same location, these observations are consistent with a mechanism in which components of the adherens junction either are diluted temporarily by the addition of membrane from surface-connected compartments during TEM or are transiently internalized during this process. VE-cadherin is constitutively internalized from cell borders (Fig. 1); however, we do not know whether it enters the same compartment as PECAM or whether it remains with it.

The purpose of the targeted delivery may be to increase the surface area of the intercellular border transiently to facilitate the passage of leukocytes. Alternatively, the purpose may not be to bring more surface area in general, but to bring more PECAM in particular, to the junction. The emigrating leukocyte expresses PECAM on its surface. This PECAM must make contact with endothelial cell PECAM for diapedesis to continue. If all of the endothelial cell PECAM at the cell borders is tied up in homophilic adhesions with the adjacent endothelial cell, however, there may be no available PECAM for the leukocyte to engage. Targeted addition of membrane bearing unengaged PECAM molecules would provide a source of PECAM for homophilic interaction with the leukocyte entering the junction. This engagement may be necessary to propagate a signal for further recruitment of this membrane farther along the junction as the leukocyte moves through. Alternatively or additionally, the accretion of free PECAM-bearing membrane in the vicinity of the migrating leukocyte might create a haptotactic gradient for the leukocyte to follow through the junction. Our identification of the role of membrane trafficking in TEM offers insight into the regulation of inflammation and might also lead to



**Figure 3** Quantification of PECAM recycling. **a**, Recycling membrane was analysed as described in Methods. Scale bar, 20  $\mu$ m. **b**, Pixel intensity was quantified from digitized images (mean  $\pm$  s.e.m. of five samples), and the results are shown graphically in arbitrary units. **c**, Monolayers treated as above were rapidly removed from dishes and

analysed by flow cytometry. Fluorescence is proportional to the amount of recycled PECAM. The ‘time zero’ specimen fluorescence includes recycling that may have taken place during removal from the plate at 37 °C.



targets for anti-inflammatory therapy as the signals that regulate this membrane movement are identified. □

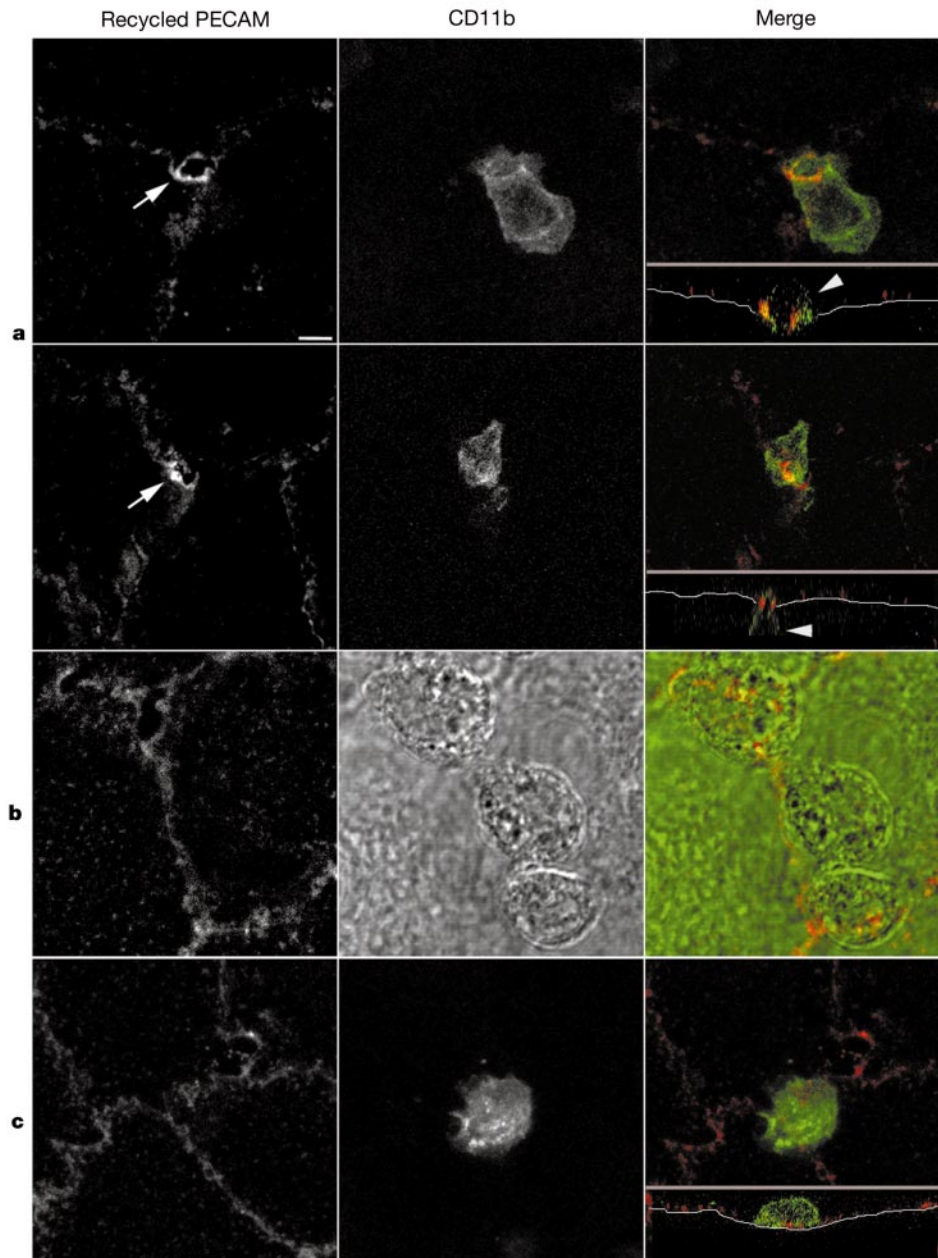
## Methods

### Cell isolation and culture

We isolated HUVEC cells and cultured them as described<sup>19</sup> on plastic or hydrated collagen gels. Transmigration experiments were carried out on confluent monolayers as described<sup>5,20</sup>, using freshly isolated peripheral blood mononuclear cells as a source of monocytes.

### Immunohistochemistry

HUVEC monolayers were washed in cold divalent cation-containing buffer and fixed in freshly prepared paraformaldehyde (2% in PBS). To visualize intracellular labelling, cells were subsequently incubated in 0.1% Triton X-100 for 30 min on ice, followed by extensive washing in cold PBS before the antibodies were added. We used the following murine monoclonal antibodies: hec7 and P1.1 against PECAM-1 (ref. 10), hec1 against VE-cadherin<sup>21</sup>, hec2 against CD99 (ref. 8), 84H10 against ICAM-1 (ref. 20) and OKM1 against CD11b (American Type Culture Collection). Routine immunohistochemistry was visualized by a Zeiss Axiophot2 microscope with a  $63 \times 1.4$  n.a. lens. Immunoelectron microscopy is described in the Supplementary Information.



**Figure 4** Recycling PECAM is involved in transendothelial migration. Recycling PECAM was visualized as in Fig. 3, but with monocytes transmigrating simultaneously. A stacked series of images through the whole thickness of the interacting cells is shown. Monocytes were preincubated in  $20 \mu\text{g ml}^{-1}$  preimmune IgG (**a**, **b**) or blocking antibodies against PECAM (**c**). Recycling PECAM was visualized with  $\text{F(ab')}_2$  fragments of rhodamine-labelled antibody against mouse IgG and monocytes were visualized with an Alexa 488-labelled monoclonal antibody against CD11b. In **a**, recycled PECAM is enriched around the monocyte as it transmigrates (arrow). Representative images of monocytes at relatively early (top) and late (bottom) stages of diapedesis are shown. Orthogonal sections through these fields (insets in merged images) show the extent of diapedesis. The plane of

the monolayer is underlined in white. Arrowheads indicate the body of the transmigrating monocyte. In **b**, THP-1 cells are adherent to the HUVEC monolayer over the cell border but cannot transmigrate. Differential interference contrast microscopy used to visualize these cells is printed green in the merged image frame. In **c**, a monocyte is tightly adherent over the junction, but cannot migrate because its PECAM is rendered nonfunctional by a blocking antibody. In the absence of a productive PECAM–PECAM interaction between monocyte and HUVEC, there is no selective recycling of internal membrane to the zone of diapedesis. Images are representative of hundreds of cells observed in  $>12$  independent experiments. Scale bar,  $5 \mu\text{m}$ .

## Membrane recycling experiments

For quantitative fluorescence microscopy, HUVEC monolayers were incubated for 1 h at 37 °C with Fab fragments of P1.1, a monoclonal antibody against domain 5 of PECAM that does not inhibit any known PECAM function<sup>10</sup>. During this time, the Fab bound to PECAM at the junction and was taken up into the subjunctional compartment. Control experiments showed that binding of Fab to PECAM was stable down to pH 3.0. We washed off unbound antibody, chilled the monolayers on ice, and added rabbit antibody against mouse IgG (RAM) for 1 h at 4 °C; these conditions were determined to saturably bind all of the available antibody that was present on PECAM molecules at the cell surface (that is, any monoclonal antibody bound to PECAM molecules that had not been sequestered, or any that had recycled back to the cell surface during the incubation). After this treatment, unbound RAM was washed away and FITC-labelled RAM was added in the cold for 1 h to allow ample time to penetrate along the endothelial cell borders. At time zero, the monolayers were rapidly warmed to 37 °C, allowing membrane trafficking to resume for various times before fixation. FITC-RAM can only bind to Fab fragments that were not previously in the junction, that is, Fab fragments on sequestered membrane that recycled to the junction or Fab fragments on membrane that now communicated with the junction. Newly synthesized PECAM molecules coming to the junction at this time would not be labelled by FITC-RAM, because they would not have had the opportunity to bind Fab fragments against PECAM.

Digitized images were made using a Leica DMIRB microscope equipped with a Princeton Instruments cooled CCD (charge-coupled device) camera driven by Image-1/ MetaMorph Imaging System software (Universal Imaging Corporation). Quantitative analysis of images was done with MetaMorph software (Universal Imaging Corporation). Briefly, after background subtraction, the total integrated intensity per field was measured for five random fields per time point. We plotted mean fluorescence intensities per field for each time point.

For flow cytometry, HUVEC monolayers were treated as above, but FITC-RAM Fab was used as the detecting agent. At specified times, cells were rapidly removed from the plates, given a brief (2-min) exposure to trypsin plus EDTA and quenched in ice-cold serum-containing medium. Pilot experiments had shown that this procedure does not affect the surface levels of immunologically detectable PECAM. We analysed cells by flow cytometry on a FACScalibur cytometer using CellQuest software (Becton Dickinson).

## Fluorescence quenching experiments

External fluorescence quenching experiments were carried out on confluent HUVEC monolayers grown on cover slip dishes. We incubated the monolayers with FITC-labelled P1.1 Fab for 1 h at 37 °C or 4 °C, and then washed the cultures and chilled them in ice-cold PBS. Digitized images of representative fields were captured using a Leica DMIRB microscope, as described above. Immediately after obtaining the image, the pH of the surrounding buffer was dropped to 5.5 by flooding the dish with 20 mM 2-(N-morpholino)-ethanesulphonic acid (MES) buffer and a second image of the same field was captured within 2 s of the change in pH (ref. 9) using the same imaging settings. Control experiments in which HUVEC monolayers were allowed to endocytose FITC-dextran into endosomal or lysosomal compartments showed that this treatment had no effect on intracellular pH.

## Recycling during TEM

Monolayers were incubated with P1.1 Fab as for the quantification of recycling experiments. After chilling and saturable binding of junctional Fab with F(ab')<sub>2</sub> fragments of RAM, freshly isolated human monocytes (or THP-1 monocyte line in one series of control experiments) were added, along with rhodamine-labelled F(ab')<sub>2</sub> fragments of RAM, at 4 °C. These monocytes had been incubated in a blocking rabbit antibody against PECAM or a preimmune IgG, and then washed extensively before their addition. After the monocytes had settled on the monolayer, the cultures were quickly warmed to 37 °C to allow a rapid and synchronous wave of TEM. After 10–15 min, TEM was stopped by rapid washing and fixation of the monolayers. The degree of TEM was analysed by a typical TEM assay<sup>5</sup>, and membrane recycling was detected by the rhodamine-conjugated antibody against mouse IgG, which would only stain the P1.1 Fab on recycling PECAM. Monocytes were identified by staining with OKM1 antibodies conjugated to Alexa-488 (Molecular Probes), according to the manufacturer's directions. We examined samples by confocal microscopy using a Zeiss LSM 510 microscope and analysed images using MetaMorph software.

Received 26 September; accepted 12 November 2002; doi:10.1038/nature01300.

- Butcher, E. C. Leukocyte-endothelial cell recognition: three (or more) steps to specificity and diversity. *Cell* **67**, 1033–1036 (1991).
- Springer, T. A. Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm. *Cell* **76**, 301–314 (1994).
- Muller, W. A. & Randolph, G. J. Migration of leukocytes across endothelium and beyond: molecules involved in the transmigration and fate of monocytes. *J. Leukoc. Biol.* **66**, 698–704 (1999).
- Muller, W. A. Leukocyte-endothelial cell interactions in the inflammatory response. *Lab. Invest.* **82**, 521–534 (2002).
- Muller, W. A., Weigl, S. A., Deng, X. & Phillips, D. M. PECAM-1 is required for transendothelial migration of leukocytes. *J. Exp. Med.* **178**, 449–460 (1993).
- Berman, M. E., Xie, Y. & Muller, W. A. Roles of platelet/endothelial cell adhesion molecule-1 (PECAM-1, CD31) in natural killer cell transendothelial migration and  $\beta$ 2 integrin activation. *J. Immunol.* **156**, 1515–1524 (1996).
- Liao, F., Ali, J., Greene, T. & Muller, W. A. Soluble domain 1 of platelet-endothelial cell adhesion molecule (PECAM) is sufficient to block transendothelial migration in vitro and in vivo. *J. Exp. Med.* **185**, 1349–1357 (1997).
- Schenkel, A. R., Mamdouh, Z., Chen, X., Liebman, R. M. & Muller, W. A. CD99 plays a major role in the migration of monocytes through endothelial junctions. *Nature Immunol.* **3**, 143–150 (2002).

- Yamashiro, D. J. & Maxfield, F. R. Kinetics of endosome acidification in mutant and wild-type chinese hamster ovary cells. *J. Cell Biol.* **105**, 2713–2721 (1987).
- Liao, F., Huynh, H. K., Eiroa, A., Greene, T., Polizzi, E. & Muller, W. A. Migration of monocytes across endothelium and passage through extracellular matrix involve separate molecular domains of PECAM-1. *J. Exp. Med.* **182**, 1337–1343 (1995).
- Feng, D., Nagy, J. A., Hipp, J., Dvorak, H. F. & Dvorak, A. M. Vesiculo-vacuolar organelles and the regulation of venule permeability to macromolecules by vascular permeability factor, histamine, and serotonin. *J. Exp. Med.* **183**, 1981–1986 (1996).
- Vasile, E., Qu, H., Dvorak, H. F. & Dvorak, A. M. Caveolae and vesiculo-vacuolar organelles in bovine capillary endothelial cells cultured with VPF/VEGF on floating Matrigel-collagen gels. *J. Histochem. Cytochem.* **47**, 159–167 (1999).
- Schmidt, A., Hannah, M. J. & Huttner, W. B. Synaptic-like microvesicles of neuroendocrine cells originate from a novel compartment that is continuous with the plasma membrane and devoid of transferrin receptor. *J. Cell Biol.* **137**, 445–458 (1997).
- Feng, D., Nagy, J. A., Pyne, K., Dvorak, H. F. & Dvorak, A. M. Neutrophils emigrate from venules by a transendothelial cell pathway in response to fMLP. *J. Exp. Med.* **187**, 903–915 (1998).
- Bamforth, S., Lightman, S. & Greenwood, J. Ultrastructural analysis of interleukin-1 $\beta$ -induced leukocyte recruitment to the rat retina. *Investig. Ophthalmol. Vis. Sci.* **38**, 25–35 (1997).
- Hammersen, F. & Hammersen, E. The ultrastructure of endothelial gap formation and leukocyte emigration. *Prog. Appl. Microcirc.* **12**, 1–34 (1987).
- Allport, J. R., Muller, W. A. & Lusinskas, F. W. Monocytes induce reversible focal changes in vascular endothelial cadherin complex during transendothelial migration under flow. *J. Cell Biol.* **148**, 203–216 (2000).
- Shaw, S. K., Bamba, P. S., Perkins, B. N. & Lusinskas, F. W. Real-time imaging of vascular endothelial-cadherin during transmigration across endothelium. *J. Immunol.* **167**, 2323–2330 (2001).
- Muller, W. A., Ratti, C. M., McDonnell, S. L. & Cohn, Z. A. A human endothelial cell-restricted, externally disposed plasmalemmal protein enriched in intercellular junctions. *J. Exp. Med.* **170**, 399–414 (1989).
- Muller, W. A. & Weigl, S. Monocyte-selective transendothelial migration: Dissection of the binding and transmigration phases by an in vitro assay. *J. Exp. Med.* **176**, 819–828 (1992).
- Ali, J., Liao, F., Martens, E. & Muller, W. A. Vascular endothelial cadherin (VE-Cadherin): cloning and role in endothelial cell-cell adhesion. *Microcirculation* **4**, 267–277 (1997).

**Supplementary Information** accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

**Acknowledgements** We thank R. Liebman for technical assistance; P. Newman for the P1.1 antibody; and P. Brennwald, T. McGraw and T. Ryan for discussions and comments on the manuscript. Supported by NIH grants (to W.A.M. and F.R.M.), a Charles H. Revson Foundation Fellowship (to Z.M.), and an Atorvastatin Research Award from Pfizer/Parke Davis (to L.P.).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to W.A.M. (e-mail: [wamuller@med.cornell.edu](mailto:wamuller@med.cornell.edu)).

# Cdc42 regulates GSK-3 $\beta$ and adenomatous polyposis coli to control cell polarity

Sandrine Etienne-Manneville & Alan Hall

MRC Laboratory for Molecular Cell Biology and Cell Biology Unit, Cancer Research UK Oncogene and Signal Transduction Group, and Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK

Cell polarity is a fundamental property of all cells. In higher eukaryotes, the small GTPase Cdc42, acting through a Par6-atypical protein kinase C (aPKC) complex, is required to establish cellular asymmetry during epithelial morphogenesis, asymmetric cell division and directed cell migration<sup>1–5</sup>. However, little is known about what lies downstream of this complex. Here we show, through the use of primary rat astrocytes in a cell migration assay, that Par6-PKC $\zeta$  interacts directly with and regulates glycogen synthase kinase-3 $\beta$  (GSK-3 $\beta$ ) to promote polarization of the centrosome and to control the direction of cell protrusion. Cdc42-dependent phosphorylation of GSK-3 $\beta$  occurs specifically at the leading edge of migrating cells, and

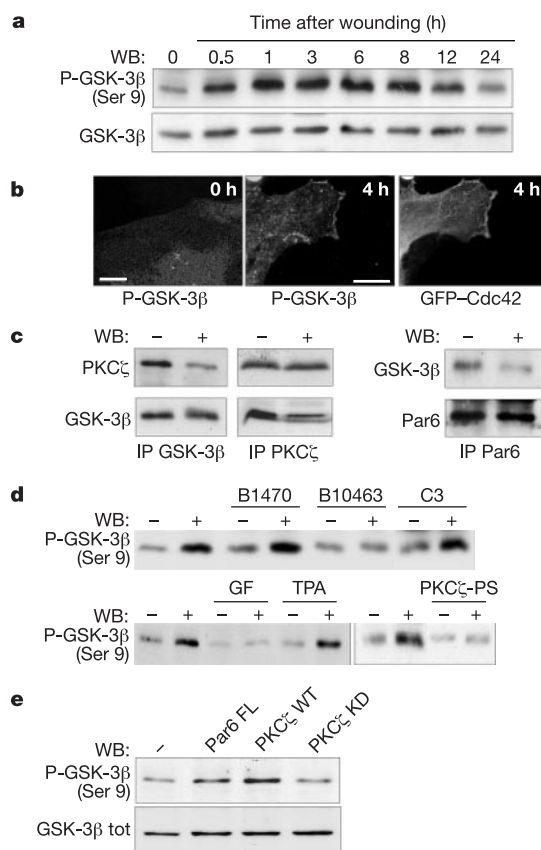
induces the interaction of adenomatous polyposis coli (Apc) protein with the plus ends of microtubules. The association of Apc with microtubules is essential for cell polarization. We conclude that Cdc42 regulates cell polarity through the spatial regulation of GSK-3 $\beta$  and Apc. This role for Apc may contribute to its tumour-suppressor activity.

Scratch-induced cell migration in astrocyte monolayers is associated with the spatially restricted activation of Cdc42, which acts through a Par6-PKC $\zeta$  complex to control cell polarity. In a search for molecules that act downstream of Par6-PKC $\zeta$ , we examined the phosphorylation state of GSK-3 $\beta$ , a protein previously implicated in the establishment of polarity<sup>6–10</sup>. The level of phosphorylation at serine 9 of GSK-3 $\beta$  increases soon after scratching a monolayer, reaches a maximum at 1 h, and lasts for at least 12 h (note that wound closure is complete by 24 h; Fig. 1a). The total levels of

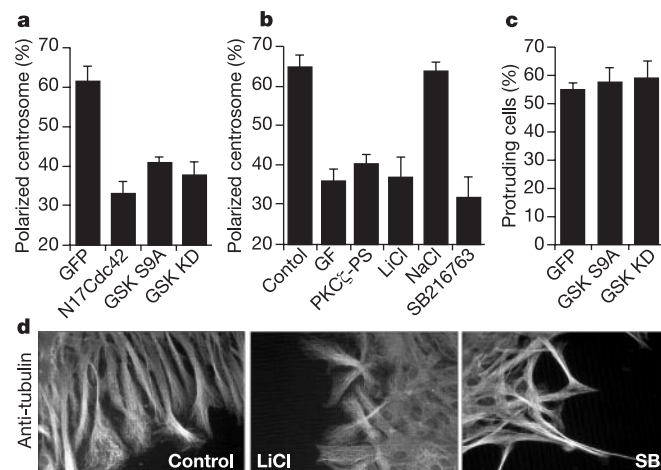
GSK-3 $\beta$  remain constant. Figure 1b shows that GSK-3 $\beta$  phosphorylation occurs preferentially at the leading edge of migrating cells, where we previously reported co-localization of Cdc42, Par6 and PKC $\zeta$ <sup>5</sup>. Ectopically expressed enhanced green fluorescent protein (EGFP)-Cdc42 localizes together with phosphorylated GSK-3 $\beta$  (Fig. 1b). To determine whether GSK-3 $\beta$  physically associates with PKC $\zeta$ <sup>11</sup>, the proteins were immunoprecipitated and analysed on western blots. We found that the two proteins can be precipitated together and exist in a complex, but after scratch-induced migration, significant dissociation occurs (Fig. 1c). Phosphorylated GSK-3 $\beta$  cannot be detected in the PKC $\zeta$  precipitate (not shown), indicating that phosphorylation of GSK-3 $\beta$  leads to its dissociation from PKC $\zeta$ . Figure 1c shows that GSK-3 $\beta$  can be precipitated with an anti-Par6 antibody, suggesting a larger complex containing GSK-3 $\beta$ , Par6 and PKC $\zeta$ .

To confirm that GSK-3 $\beta$  phosphorylation is dependent on the Cdc42-Par6-PKC $\zeta$  pathway, cells were pre-treated with toxin B10463 (inhibits Cdc42, Rac, Rho), toxin B1470 (inhibits Rac, Ral, Rap1, R-ras) and C3 transferase (inactivates Rho). Only the toxin that inhibits Cdc42 prevents GSK-3 $\beta$  phosphorylation (Fig. 1d). Inhibition of all PKC isoforms with GF109203X (Fig. 1d) or RO 31-8220 (not shown), or inhibition of PKC $\zeta$  with a cell-permeable PKC $\zeta$ -specific pseudo-substrate (Fig. 1d), prevents GSK-3 $\beta$  phosphorylation, whereas depletion of conventional or novel PKCs by 12-O-tetradecanoylphorbol-13-acetate (TPA) treatment does not (Fig. 1d). Furthermore, inhibition of phosphatidylinositol-3-OH kinase (PI(3)K) with wortmannin or LY 294002 has no effect (not shown), indicating that Akt/PKB is not involved. Finally, transfection of COS cells with Par6, which activates PKC $\zeta$ , or with wild-type (but not kinase-dead) PKC $\zeta$  induces phosphorylation of GSK-3 $\beta$  on serine 9 (Fig. 1e).

Phosphorylation of GSK-3 at serine 9 inhibits its catalytic activity<sup>12,13</sup>. To explore the significance of this, GSK-3 S9A, a non-phosphorylatable, constitutively activated mutant was microinjected into leading-edge cells. The establishment of polarity was visualized by the reorientation of the centrosome to face the direction of migration, and by the formation of cell protrusions perpendicular to the scratch<sup>5</sup>. GSK-3 S9A is almost as effective as dominant-negative Cdc42 (N17Cdc42) in blocking the reorientation of the centrosome (Fig. 2a). Expression of a kinase-dead



**Figure 1** Glycogen synthase kinase-3 $\beta$  (GSK-3 $\beta$ ) is phosphorylated downstream of Cdc42 and protein kinase C $\zeta$  (PKC $\zeta$ ) during astrocyte migration. **a**, Astrocyte monolayers were scratched and were incubated for the indicated times. Phosphorylated (Ser 9) and total GSK-3 $\beta$  were analysed by western blotting (WB). **b**, Cell monolayers were scratched and immediately microinjected with green fluorescent protein (GFP)-Cdc42. Cells were fixed 0 or 4 h later and stained with anti-phospho-GSK-3 (Ser 9) antibody (left and middle panels). GFP-Cdc42 (right panel) localizes with phospho-GSK-3 (Ser 9) at the leading edge. Scale bar, 10  $\mu$ m. **c**, Astrocytes were lysed before (-) or 1 h after (+) scratching. Endogenous GSK-3 $\beta$  or PKC $\zeta$  were immunoprecipitated (IP) and analysed on western blots with anti-PKC $\zeta$  or anti-GSK-3 $\beta$  antibodies, respectively. Par6 was immunoprecipitated and analysed with anti-Par6 or anti-GSK-3 $\beta$  antibodies. **d**, Astrocytes were pre-treated with toxin B 1470 (3 h, 10  $\mu$ g ml<sup>-1</sup>), toxin B 10463 (3 h, 1  $\mu$ g ml<sup>-1</sup>), C3 toxin (overnight, 5  $\mu$ g ml<sup>-1</sup>), GF109203X (20  $\mu$ M, 1 h), PKC $\zeta$  pseudo-substrate (10  $\mu$ M, 1 h) or 12-O-tetradecanoylphorbol-13-acetate (TPA) (160 nM, overnight), scratched in the presence of the inhibitors, and incubated for 1 h. GSK-3 $\beta$  phosphorylation was visualized as above. **e**, COS cells were transfected and lysed after 48 h, and phosphorylated (Ser 9) and total GSK-3 $\beta$  visualized. Each experiment was repeated three times. FL, full length; KD, kinase-dead; WT, wild type.

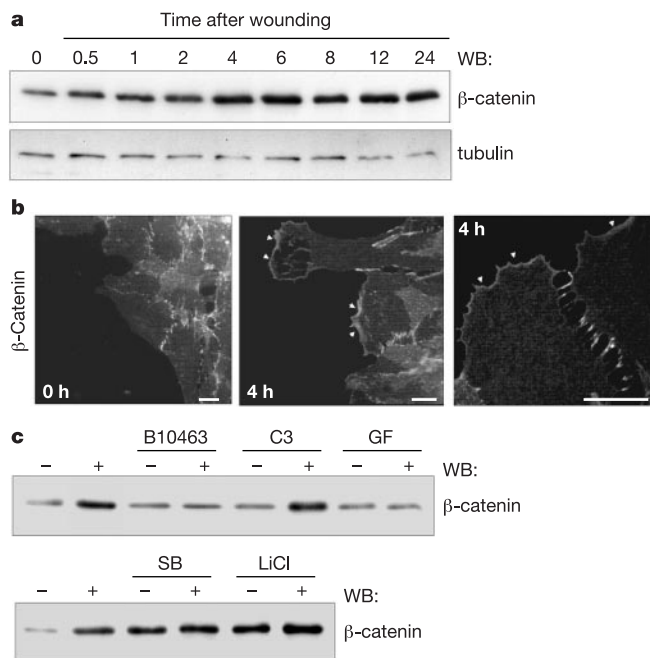


**Figure 2** Spatially localized inhibition of GSK-3 is required to establish cell polarity. **a**, **b**, After scratching a monolayer, leading-edge cells were microinjected with the indicated constructs (**a**), or incubated with inhibitors (**b**). Centrosome polarization was determined 8 h later. **c**, Cells at the leading edge were microinjected, and the number of expressing cells with protrusions was determined 8 h later. **d**, Cells were scratched, incubated with LiCl (20  $\mu$ M) or SB216763 (20  $\mu$ M), and 16 h later they were fixed and stained with an anti-tubulin antibody.



version of GSK-3 $\beta$  also inhibits centrosome reorientation (Fig. 2a). This is reminiscent of our previous observations that dominant-negative or constitutively activated Cdc42 inhibit polarity, and points to the importance of localization of signalling activities<sup>5</sup>. To visualize the effects of global GSK-3 inhibition, cells were pre-treated with two chemically distinct inhibitors, LiCl and SB216763, both of which block centrosome polarity (Fig. 2b). Close inspection (Fig. 2c, d) reveals that in the presence of GSK-3 inhibitors, cell protrusions still form but are randomly oriented.

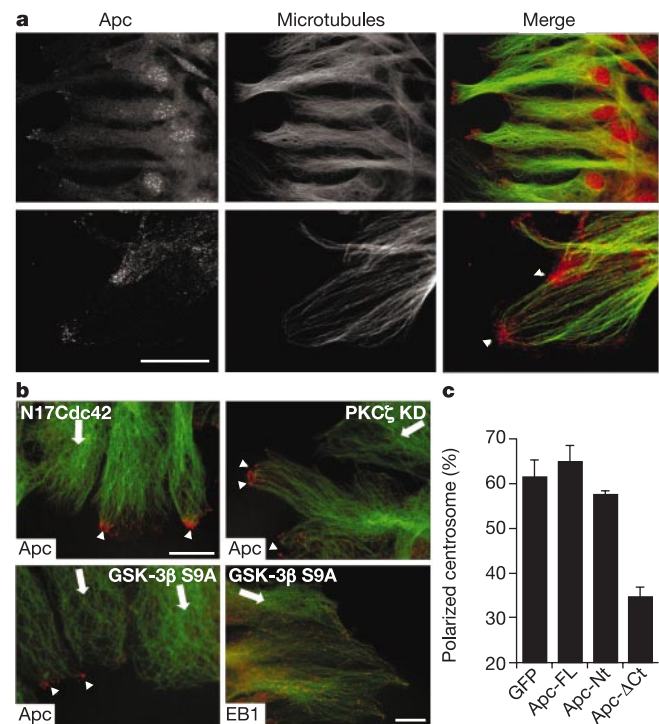
GSK-3 is a component of numerous signal transduction pathways. In the Wnt pathway, GSK-3 is part of an axin–Apc– $\beta$ -catenin complex, and Wnt induces inactivation of GSK-3, stabilization of  $\beta$ -catenin and induction of gene transcription<sup>14,15</sup>. The loss of APC activity in human cancers also leads to  $\beta$ -catenin stabilization<sup>16</sup>. To determine the effects of localized GSK-3 inhibition in migrating cells, lysates from resting and scratched astrocyte monolayers were analysed on western blots. Thirty minutes after scratch-induced migration, stabilization of  $\beta$ -catenin is observed (Fig. 3a). This persists for at least 24 h and is dependent on Cdc42 and PKC $\zeta$  (Fig. 3a, c). Figure 3b reveals that  $\beta$ -catenin accumulates at the leading edge of migrating cells and is still visible at cell–cell contacts, which are maintained during the migration assay. There is no detectable increase in nuclear  $\beta$ -catenin or inhibition of astrocyte polarization on treatment with cycloheximide or actinomycin D. Accumulation of  $\beta$ -catenin at the leading edge of cells is abolished by expression of N17Cdc42, kinase-dead PKC $\zeta$  or GSK-3 $\beta$  S9A (data not shown). It can also be disrupted with cytochalasin D (not shown), but, in this case, cell polarity is not affected<sup>5</sup>. The data indicate that the accumulation of  $\beta$ -catenin at the leading edge reflects the localized inhibition of GSK-3, but is not required for cell polarity. It may, however, contribute other activities to the process of cell migration.



**Figure 3**  $\beta$ -Catenin is stabilized and localized at the leading edge of migrating cells. **a**, Scratched monolayers were incubated for different times, and  $\beta$ -catenin levels were assessed by western blot. An anti-tubulin antibody was used as a loading control. **b**, Cells were fixed and stained with anti- $\beta$ -catenin antibody 4 h after scratching. Arrowheads show  $\beta$ -catenin localized at the leading edge of migrating cells. Scale bar, 10  $\mu$ m. **c**,  $\beta$ -Catenin levels in astrocytes pre-treated with toxin B 10463 (3 h, 1  $\mu$ g ml<sup>-1</sup>), C3 toxin (overnight, 5  $\mu$ g ml<sup>-1</sup>), GF109203X (20  $\mu$ M, 1 h), LiCl (1 h, 20 mM) or SB216763 (1 h, 20  $\mu$ M), scratched in the presence of the inhibitors, and incubated for 6 h before lysis.

The tumour-suppressor gene product Apc is phosphorylated by GSK-3 (refs 17, 18). Apc participates in the destabilization of  $\beta$ -catenin, but can also associate with the plus ends of microtubules and regulate microtubule dynamics<sup>19,20</sup>. Figure 4a, b shows that 2–4 h after scratch-induced migration, Apc becomes associated with the plus ends of microtubules specifically at the leading edge, and this association is dependent on Cdc42, PKC $\zeta$  and phosphorylation of GSK-3 at serine 9. Through residues located in the carboxy-terminal region, Apc can associate with microtubules directly, or by means of the protein EB1 (refs 21, 22). EB1 also localizes at the plus ends of microtubules, but unlike Apc, this is independent of Cdc42, PKC $\zeta$  (data not shown) and GSK-3 S9A (Fig. 4b). To determine whether the association of Apc with microtubules is required for polarization, Apc lacking the microtubule- and EB1-binding sites (Apc- $\Delta$ Ct) was expressed in leading-edge cells. This inhibits centrosome reorientation, whereas full-length Apc or the amino-terminal domain of Apc does not (Fig. 4c). None of the Apc constructs inhibits protrusion formation (data not shown).

We conclude that Cdc42 activates Par6–PKC $\zeta$ , leading to phosphorylation and inactivation of GSK-3 $\beta$  at the leading edge of migrating astrocytes. Inactivation of GSK-3 $\beta$  may affect several microtubule-associated proteins<sup>23,24</sup>, but a key consequence is the spatially restricted association of Apc with the plus ends of micro-



**Figure 4** The association of Apc with microtubules is regulated by the Cdc42–PKC $\zeta$ –GSK-3 pathway. **a**, Astrocyte monolayers were scratched and fixed 8 h later. Cells were stained with anti-tubulin (green) and anti-Apc (red) antibody. Scale bar, 10  $\mu$ m. Apc localization at the plus ends of microtubules was confirmed by expression of a GFP-tagged protein (data not shown), but no accumulation of GFP–Apc was visible in the nucleus, indicating that the nuclear staining seen here may be an artefact. **b**, Astrocyte monolayers were scratched and leading-edge cells were immediately microinjected with the indicated constructs. Four hours later, cells were fixed and stained with anti-tubulin (green), anti-Apc or anti-EB1 (lower-right panel) (red) antibody and an antibody to detect the expressed proteins. Expressing cells are indicated with a white arrow; arrowheads indicate Apc association with the microtubule plus-ends in the non-injected cells. Scale bar, 10  $\mu$ m. **c**, Astrocytes were microinjected immediately after scratching with the indicated constructs (*Xenopus* Apc full length (FL), amino acids 1–2830; APC-Nt, amino acids 1–1035; APC- $\Delta$ Ct, amino acids 1–1962). Centrosome polarization was determined 8 h after wounding.

tubules, which is essential for establishing cell polarity. The role of Apc is unclear; it could directly affect microtubule dynamics<sup>25,26</sup>, or—perhaps through EB1—interact with the dynactin–dynein complex, which is required for centrosome reorientation<sup>5,27,28</sup>. These observations link together two important signalling pathways that have been independently implicated in controlling cell polarity during the development of multicellular organisms: Cdc42–Par proteins on one hand<sup>2</sup>, and GSK-3– $\beta$ -catenin–Apc on the other<sup>10</sup>. The essential role described here for Apc in regulating the polarity of migrating cells may be an important aspect of its functional contribution to human cancer. □

## Methods

### Materials

We used the following reagents: anti- $\alpha$ -tubulin (Sigma), anti-GSK-3 $\beta$ , anti- $\beta$ -catenin, anti-EB1 (all Transduction Labs), anti-PKC $\zeta$ , anti-Cdc42 (both Santa Cruz Biotechnology), anti-pericentrin (BabCO), anti-phospho-GSK-3 $\beta$  (Ser 9) (Biosource), anti-Apc (I. Näthke) and anti-Par6C (amino acids 2–16; P. Aspenström), GF109203X, RO 31-8220 and SB216763 (all Calbiochem), PKC $\zeta$  pseudo-substrate (Biosource), toxins B10463 and B1470 (C. von Eichel-Streiber), GSK-3 constructs (V. M. Lee and R. Kypta) and Apc constructs (B. M. Gumbiner). GTPases, Par6 and PKC $\zeta$  constructs have been described previously<sup>5</sup>.

### Cell culture and scratch-induced migration

Primary astrocyte monolayers and scratch assays have been described previously<sup>5</sup>. Individual wounds around 300  $\mu$ m wide were made with a microinjection needle, and wound closure occurred 16–24 h later. For biochemical analysis, multiple wounds were made with an eight-channel pipette (with 2- $\mu$ l tips) scratched several times across a 90-mm dish. Centrosomes were localized using anti-pericentrin antibody<sup>5</sup>, and those within the quadrant facing the wound were scored as positive. We examined 300 cells for each condition. To assess polarized morphology, astrocytes were microinjected with pEGFP as a control or with the indicated DNA constructs and biotin-dextran, and were fixed 8 h later. Cells expressing these constructs were scored as protruding when their length was at least four times their width.

### Immunoprecipitation

Cells were washed with ice-cold PBS containing 1 mM orthovanadate and lysed at 4 °C in buffer (10 mM Tris-HCl, pH 7.5, 140 mM NaCl, 1 mM orthovanadate, 1% Nonidet-P40, 2 mM phenylmethylsulphonyl fluoride, 5 mM EDTA, 20  $\mu$ g ml<sup>-1</sup> aprotinin, 20  $\mu$ g ml<sup>-1</sup> leupeptin). Nuclei were discarded after centrifugation at 10,000g for 10 min. Lysates were incubated for 2 h at 4 °C with specific antibodies and protein G Sepharose, and immunoprecipitates were collected by centrifugation and were analysed on 8% SDS–polyacrylamide gel electrophoresis gels.

Received 25 November 2002; accepted 14 January 2003; doi:10.1038/nature01423.

Published online 29 January 2003.

- Etienne-Manneville, S. & Hall, S. Rho GTPases in cell biology. *Nature* **420**, 629–635 (2002).
- Ohno, S. Intercellular junctions and cellular polarity: the PAR–aPKC complex, a conserved core cassette playing fundamental roles in cell polarity. *Curr. Opin. Cell Biol.* **13**, 641–648 (2001).
- Gotta, M., Abraham, M. C. & Ahninger, J. CDC-42 controls early cell polarity and spindle orientation in *C. elegans*. *Curr. Biol.* **11**, 482–488 (2001).
- Kay, A. J. & Hunter, C. P. CDC-42 regulates PAR protein localization and function to control cellular and embryonic polarity in *C. elegans*. *Curr. Biol.* **11**, 474–481 (2001).
- Etienne-Manneville, S. & Hall, A. Integrin-mediated Cdc42 activation controls cell polarity in migrating astrocytes through PKC $\zeta$ . *Cell* **106**, 489–498 (2001).
- Dominguez, I., Itoh, K. & Sokol, S. Y. Role of glycogen synthase kinase 3 $\beta$  as a negative regulator of dorsoventral axis formation in *Xenopus* embryos. *Proc. Natl Acad. Sci. USA* **92**, 8498–8502 (1995).
- Emily-Fenouil, F., Ghigliione, C., Lhomond, G., Lepage, T. & Gache, C. GSK3 $\beta$ /shaggy mediates patterning along the animal-vegetal axis of the sea urchin embryo. *Development* **125**, 2489–2498 (1998).
- He, X., Saint-Jeannet, J.-P., Woodgett, J. R., Varmus, H. E. & Dawid, I. Glycogen synthase kinase-3 and dorsoventral patterning in *Xenopus* embryos. *Nature* **374**, 617–622 (1995).
- Pierce, S. B. & Kimelman, D. Regulation of Spemann organizer formation by intracellular Xgsk-3. *Development* **121**, 755–765 (1995).
- Ferkey, D. M. & Kimelman, D. GSK-3: new thoughts on an old enzyme. *Dev. Biol.* **225**, 471–479 (2000).
- Oriente, F. *et al.* Insulin receptor substrate-2 phosphorylation is necessary for protein kinase C $\zeta$  activation by insulin in L6hIR cells. *J. Biol. Chem.* **276**, 37109–37119 (2001).
- Harwood, J. A. Regulation of GSK-3: a cellular multiprocessor. *Cell* **105**, 821–824 (2001).
- Troussard, A. A., Tan, C., Yoganathan, T. N. & Dedhar, S. Cell–extracellular matrix interactions stimulate the AP-1 transcription factor in an integrin-linked kinase- and glycogen synthase kinase 3-dependent manner. *Mol. Cell Biol.* **19**, 7420–7427 (1999).
- Li, L. *et al.* Axin and Frat1 interact with Dvl and GSK, bridging Dvl to GSK in the Wnt-mediated regulation of LEF-1. *EMBO J.* **18**, 4233–4240 (1999).
- Moon, R. T., Bowerman, B., Boutros, M. & Perrimon, N. The promise and perils of Wnt signaling through  $\beta$ -catenin. *Science* **296**, 1644–1646 (2002).
- Polakis, P. Wnt signaling and cancer. *Genes Dev.* **14**, 1837–1851 (2000).
- Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. Regulation of intracellular  $\beta$ -catenin levels by the adenomatous polyposis coli (APC) tumour-suppressor protein. *Proc. Natl Acad. Sci. USA* **92**, 3046–3050 (1995).
- Rubinfeld, B. *et al.* Binding of GSK3 $\beta$  to the APC- $\beta$ -catenin complex and regulation of complex assembly. *Science* **272**, 1023–1026 (1996).

- Bienz, M. The subcellular destinations of APC proteins. *Nature Rev. Mol. Cell Biol.* **3**, 328–338 (2002).
- Mogensen, M. M., Tucker, J. B., Mackie, J. B., Prescott, A. R. & Nathke, I. S. The adenomatous polyposis coli protein unambiguously localizes to microtubule plus ends and is involved in establishing parallel arrays of microtubule bundles in highly polarized epithelial cells. *J. Cell Biol.* **157**, 1041–1048 (2002).
- Su, L. K. *et al.* APC binds to the novel protein EB1. *Cancer Res.* **55**, 2972–2977 (1995).
- Barth, A. I. M., Siemers, K. A. & Nelson, W. J. Dissecting interactions between EB1, microtubules and APC in cortical clusters at the plasma membrane. *J. Cell Sci.* **115**, 1583–1590 (2002).
- Wagner, U., Utton, M., Gallo, J.-M. & Müller, C. C. J. Cellular phosphorylation of Tau by GSK-3 $\beta$  influences tau binding to microtubules and microtubule organisation. *J. Cell Sci.* **109**, 1537–1543 (1996).
- Lucas, F. R., Goold, R. G., Gordon-Weeks, P. R. & Salinas, P. C. Inhibition of GSK-3 $\beta$  leading to the loss of phosphorylated MAP-1B is an early event in axonal remodelling induced by WNT-7a or lithium. *J. Cell Sci.* **111**, 1351–1361 (1998).
- Nakamura, M., Zhou, X. Z. & Lu, K. P. Critical role for the EB1 and APC interaction in the regulation of microtubule polymerization. *Curr. Biol.* **11**, 1062–1067 (2001).
- Zumbrunn, J., Kinoshita, K., Hyman, A. A. & Nathke, I. S. Binding of the adenomatous polyposis coli protein to microtubules increases microtubule stability and is regulated by GSK3 $\beta$  phosphorylation. *Curr. Biol.* **11**, 44–49 (2001).
- Palazzo, A. F. *et al.* Cdc42, dynein, and dynactin regulate MTOC reorientation independent of Rho-regulated microtubule stabilization. *Curr. Biol.* **11**, 1536–1541 (2001).
- Berrueta, L., Tirnauer, J. S., Schuyler, S. C., Pellman, D. & Bierer, B. E. The APC-associated protein EB1 associates with components of the dynactin complex and cytoplasmic dynein intermediate chain. *Curr. Biol.* **9**, 425–428 (1999).

**Acknowledgements** This work was supported by a Cancer Research UK programme grant, the Medical Research Council and by an EMBO Long-Term Fellowship (S.E.-M.). We thank S. Martin, V. M. Lee, R. Kypta, B. M. Gumbiner, P. Aspenström, I. Näthke and C. von Eichel-Streiber for plasmids and reagents.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to A.H. (e-mail: alan.hall@ucl.ac.uk).

## Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab

Hyun-Soo Cho<sup>\*†</sup>, Karen Mason<sup>‡</sup>, Kasra X. Ramyar<sup>\*</sup>, Ann Marie Stanley<sup>\*</sup>, Sandra B. Gabelli<sup>\*</sup>, Dan W. Denney Jr<sup>‡</sup> & Daniel J. Leahy<sup>\*†</sup>

<sup>\*</sup> Department of Biophysics and Biophysical Chemistry, and <sup>†</sup> Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA

<sup>‡</sup> Genitope Corporation, 525 Penobscot Drive, Redwood City, California 94063, USA

HER2 (also known as Neu, ErbB2) is a member of the epidermal growth factor receptor (EGFR; also known as ErbB) family of receptor tyrosine kinases, which in humans includes HER1 (EGFR, ERBB1), HER2, HER3 (ERBB3) and HER4 (ERBB4)<sup>1</sup>. ErbB receptors are essential mediators of cell proliferation and differentiation in the developing embryo and in adult tissues<sup>2</sup>, and their inappropriate activation is associated with the development and severity of many cancers<sup>3</sup>. Overexpression of HER2 is found in 20–30% of human breast cancers, and correlates with more aggressive tumours and a poorer prognosis<sup>4</sup>. Anticancer therapies targeting ErbB receptors have shown promise, and a monoclonal antibody against HER2, Herceptin (also known as trastuzumab), is currently in use as a treatment for breast cancer<sup>5</sup>. Here we report crystal structures of the entire extracellular regions of rat HER2 at 2.4 Å and human HER2 complexed with the Herceptin antigen-binding fragment (Fab) at 2.5 Å. These structures reveal a fixed conformation for HER2 that resembles a ligand-activated state, and show HER2 poised to interact with other ErbB receptors in the absence of direct ligand

**binding. Herceptin binds to the juxtamembrane region of HER2, identifying this site as a target for anticancer therapies.**

ErbB receptors comprise an extracellular region of about 630 amino acids that contains four domains (I/L1, II/CR1, III/L2 and IV/CR2) arranged as a tandem repeat of a two-domain unit, a single membrane-spanning region and a cytoplasmic tyrosine kinase<sup>6</sup>. Binding of ligand to the extracellular region induces receptor dimerization and activation of the cytoplasmic kinase, which in turn leads to autophosphorylation and initiation of downstream signalling events<sup>1,7</sup>. Over 11 different EGF-like ErbB ligands have been identified<sup>8</sup>. Many of these ligands activate homodimeric receptor complexes, but most, if not all, also signal through heteromeric receptor complexes that include HER2. Despite functioning as a universal ErbB co-receptor, HER2 is the only ErbB receptor for which a high-affinity ligand has not been found. HER2 is also unusual among ErbB receptors in its ability to transform cells in a ligand-independent manner when overexpressed<sup>3,9</sup>.

Recent crystal structures of extracellular regions of HER1 (sHER1) complexed with either EGF<sup>10</sup> or transforming growth factor- $\alpha$  (TGF- $\alpha$ )<sup>11</sup> reveal ligand bound to domains I and III, and a long, finger-like projection from domain II mediating an inter-receptor dimer. By contrast, crystal structures of non-activated forms of sHER1 (ref. 12) and sHER3 (ref. 13) reveal this finger-like projection from domain II making an intramolecular contact with a pocket on domain IV. Together these observations suggest a model for receptor activation in which unliganded receptors exist in an autoinhibited or 'closed' form that undergoes substantial domain rearrangement to an 'open' form on ligand binding. This rearrangement brings domains I and III close together, breaking the domain II–IV interaction and freeing the domain II projection to participate in receptor dimerization and initiation of signal transduction.

To provide a molecular basis for understanding the role of HER2 in signalling and cancer, we have determined the crystal structure of rat sHER2 at 2.4 Å resolution and human sHER2 complexed with the Fab fragment of the anticancer monoclonal antibody Herceptin at 2.5 Å resolution. Both rat and human sHER2 were expressed in mammalian cells and enzymatically deglycosylated before crystallization. The crystal structures were solved by molecular replacement using the A5B7 Fab structure<sup>14</sup> and the individual domains of sHER3 (ref. 13) as search models. The  $R_{\text{cryst}}/R_{\text{free}}$  values for rat sHER2 and the human sHER2–Herceptin complex are 0.226/0.282 and 0.223/0.286, respectively. Complete data collection and refinement statistics are available as Supplementary Information.

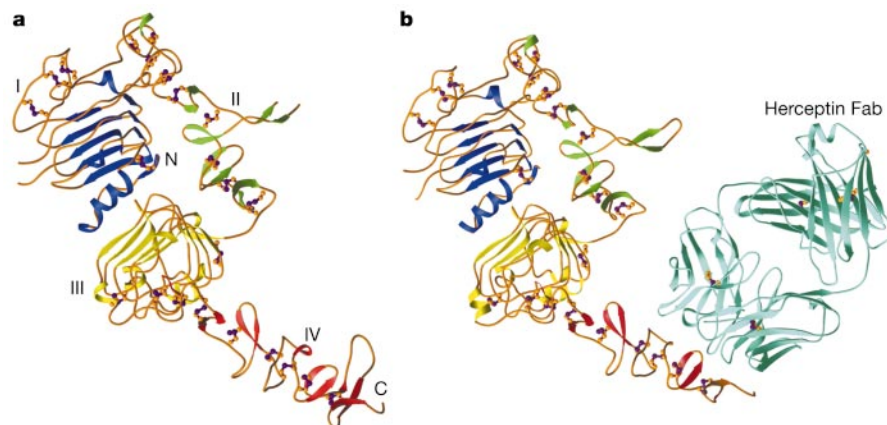
The structures of rat and human sHER2 (Fig. 1) superimpose

well. A total of 448 residues between amino acids 1 and 542 superimpose with a root mean square deviation (r.m.s.d.) of C $\alpha$  positions of 0.99 Å, with the principal variation being an approximately 7° change in the angle between domains III and IV. The appearance of such similar structures in different crystal environments indicates that the interdomain relationships in sHER2 are relatively fixed.

Comparison of the sHER2 structures with the sHER3 (ref. 13) and sHER1 (ref. 12) structures indicates that domains I and II adopt a relatively fixed interdomain relationship, as do domains III and IV, but that the relative orientation of these domain pairs varies widely (Fig. 2). Some spine-like bending of the cysteine-rich domains is also observed. The relative orientations of domain II to I and domain IV to III appear stabilized in part by conserved tryptophan residues that extend from domain II (Trp 183) and domain IV (Trp 499) to contact conserved tryptophan residues in the hydrophobic cores of domain I (Trp 147) and domain III (Trp 460), respectively. Conformational change in ErbB receptors thus seems to involve rigid-body movements of the domain I–II pair relative to the domain III–IV pair about a pivot between domains II and III. The differences in main-chain conformation that account for this pivot occur within residues 316–323 of sHER2 (307–314 of sHER3), with the largest rotation occurring about Val 320 (Ala 311 in sHER3).

The domain II–IV contact that restricts the domain arrangement in non-activated sHER1 and sHER3 is absent in sHER2 (Figs 1 and 2c), and the orientation of domains I/II relative to domains III/IV in sHER2 appears to be determined chiefly by a large interface between domains I and III. Failure to conserve two key residues in the domain IV contact region (Gly 563 and His 565 of HER3 are replaced with Pro and Phe, respectively, in HER2) may explain the absence of the domain II–IV contact in HER2 (ref. 13). Several features of the HER2 domain I–III interface, which is made up of a core of hydrophobic residues surrounded by hydrophilic contacts, suggest that it is stable; in addition to being preserved in two different crystal lattices, it buries 1,167 Å<sup>2</sup> of surface area and gives a very high shape complementarity parameter (0.75; ref. 15). Orthogonal views of this interface and an alignment of the sequences composing it are available as Supplementary Information.

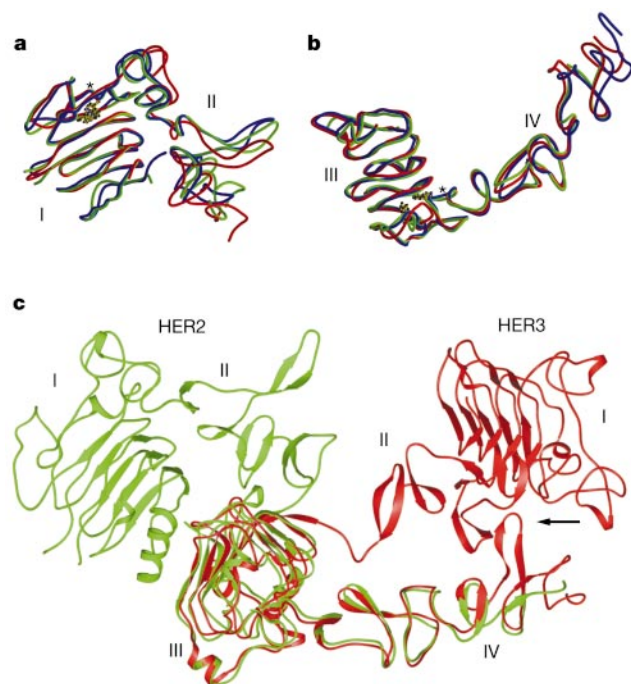
The proximity and relative orientation of domains I and III in sHER2 are very similar to the domain arrangement observed in the structures of sHER1 complexed with either TGF- $\alpha$ <sup>11</sup> or EGF<sup>10</sup> (Fig. 3). Recent biochemical evidence implicates domain IV in mediating inter-receptor dimers<sup>16</sup>, but domain IV is either not present<sup>11</sup> or not visible<sup>10</sup> in the ligand-activated structures of



**Figure 1** The structure of rat and human sHER2. **a**, Ribbon diagram of rat sHER2. Domains I (blue), II (green), III (yellow) and IV (red), and the amino and carboxy termini, are indicated. Disulphide bonds are shown in purple and gold. **b**, Ribbon diagram of the

human sHER2 (coloured as in **a**) and Herceptin Fab (cyan) complex. Figures 1 and 2c were made using the program Ribbons<sup>29</sup>.



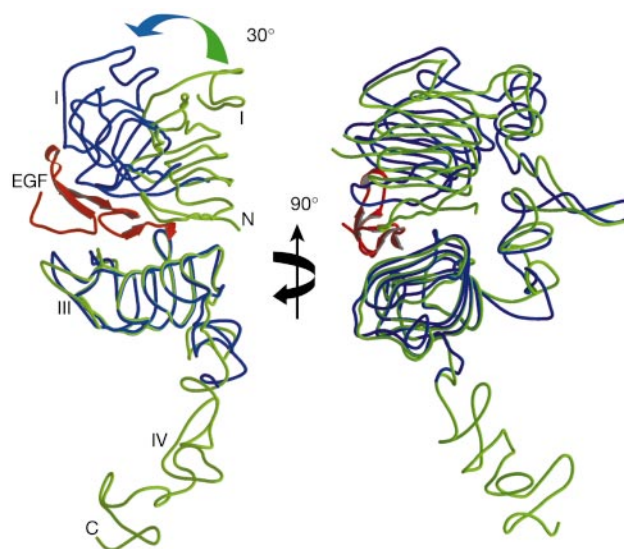


**Figure 2** Superposition of different ErbB receptors. **a**, Superposition of domains I and II of sHER1 (ref. 12; blue), sHER2 (green) and sHER3 (ref. 13; red). Ball-and-stick models of conserved tryptophan residues from each receptor are also shown and are indicated by an asterisk. **b**, Superposition of domains III and IV of sHER1, sHER2 and sHER3 coloured as in **a**. **c**, Human sHER2 (green) and sHER3 (red) after superposition of domains III and IV. Individual domains are numbered and the 'snap-like' contact between domains II and IV of sHER3 is indicated by an arrow. Figures 2a, b and 3 are made with the program MOLSCRIPT<sup>30</sup>.

sHER1. Conservation of the relative orientation of domains III and IV in ErbB receptors (Fig. 2b) allows positioning of domain IV in the ligand-induced EGFR dimer by superposing domain III of the sHER2 domain III–IV pair with domain III of the EGFR dimer. This superposition indicates that the carboxy termini of domain IV are brought into close proximity in this dimer (see Supplementary Information), which is consistent with a model in which signalling is initiated by juxtaposition of the receptor transmembrane regions, presumably in a favoured orientation.

The exposed domain II loop that mediates inter-receptor dimers of sHER1 is also exposed in sHER2 but does not mediate a similar HER2 dimer in either crystal form, indicating a dimerization constant greater than about 10 mM. The EGFR dimer is mediated entirely by residues from domain II (refs 10, 11), and the only residue at the dimer interface not conserved between EGFR and HER2 is Arg 285 of EGFR, which is Leu in HER2. Mutation of Arg 285 to Ser reduces ligand-induced phosphorylation of EGFR<sup>10</sup>, suggesting that Leu at this position in HER2 is likely to be at least partly responsible for the absence of sHER2 homodimers. Given the symmetry of the inter-receptor dimer, mutations affecting dimerization would be doubly harmful to homodimers relative to heterodimers, which may explain the apparent ability of HER2 to participate in heterodimeric complexes. It is also possible that HER2 interacts with other ErbB receptors either indirectly or through different contact regions, but the highly conserved contact residues in the domain II loop make these alternatives seem less probable.

The constitutive open structure of HER2 helps to explain several of its singular properties. The absence of a high-affinity ligand for HER2 can be explained by its fixed open structure, which does not require ligand binding to release an autoinhibited conformation. By adopting a fixed conformation, HER2 must be intrinsically capable

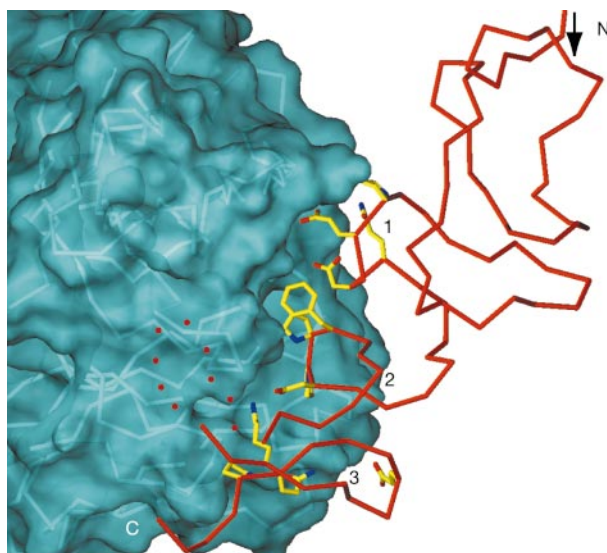


**Figure 3** Relationship of sHER2 to ligand-activated sHER1. Orthogonal views of sHER2 (green) and sHER1 (blue) with EGF (red) bound<sup>10</sup> after superposition of domain III. Domain II has been omitted from the figure on the left for clarity.

of interacting with any binding partners that become available, explaining its readiness to serve as a co-receptor with each of the other ErbB receptors in the absence of direct ligand binding. Overexpression of HER2 in tissue culture cells leads to HER2 activation and a transformed phenotype, but overexpression of HER1 or HER3 is not transforming unless ligand is added<sup>3,9</sup>. Maintenance of HER1 and HER3 in the closed conformation in the absence of ligand is clearly a barrier to autoactivation. Absence of this autoinhibited conformation for HER2 eliminates this barrier, which accounts in part for its transforming potential when overexpressed.

The anti-HER2 monoclonal antibody Herceptin has an antiproliferative effect on cells transformed by overexpression of HER2, and is an effective treatment for human breast cancer<sup>5</sup>. The mechanism of Herceptin action is thought to involve many processes, including antibody-mediated cytotoxicity and blocking receptor aggregation, but stimulation of HER2 endocytosis and removal of HER2 from the cell surface seems to be a principal element of its effectiveness<sup>17,18</sup>. Monovalent Herceptin Fab fragments do not possess the antiproliferative properties of the whole Herceptin antibody, suggesting that either receptor crosslinking is an important feature of Herceptin activity or that divalence is needed to increase its apparent affinity. Herceptin also inhibits activation of HER2 through cleavage of the juxtamembrane region of the extracellular domain, a process that normally occurs at a low level but increases when HER2 is overexpressed<sup>19</sup>.

Herceptin binds HER2 on the C-terminal portion of domain IV at a site that encompasses the binding pocket for the extended domain II loop in inactive states of HER3 and HER1 (Figs 1b and 4). This interaction buries 1,350 Å<sup>2</sup> of surface area over a long groove and possesses an unusually high shape complementarity (0.76; ref. 15) for antigen–antibody interactions. The interaction is mediated by three regions of HER2; loops formed by residues 557–561 and 570–573, and the base of a loop formed by residues 593–603. Interactions formed by the first and third loop are primarily electrostatic, whereas the second loop makes mostly hydrophobic contacts in a pocket formed by the CDR3 loops of heavy and light chains of the antibody. Binding of Herceptin to the juxtamembrane region may sufficiently crosslink the HER2 receptor to facilitate engagement of the endocytotic machinery while avoiding undue kinase activation by providing a steric barrier



**Figure 4** The Herceptin-binding site. The backbone of sHER2 is shown in red with side chains of residues contacting the Herceptin Fab added. The surface of the Herceptin Fab fragment is shown in cyan. The N and C termini of the displayed region of HER2 are indicated, as are the three loop regions of HER2 (residues 557–561 (loop 1), 570–573 (loop 2) and 593–603 (loop 3)) that contact the Herceptin Fab fragment. An eight-residue loop not visible in the structure is indicated by red dots. This figure was made with the program SPOCK.

to direct interaction of the transmembrane regions, which may be required for optimal signalling<sup>20</sup>. Binding at this site will also sterically block activation of HER2 by proteolytic cleavage of its extracellular region, and may disrupt interactions between HER2 and other proteins.

The structures of sHER2 and the Herceptin Fab fragment provide a basis for the design of new therapeutics to target HER2. Antibodies raised specifically against the juxtamembrane region of ErbB receptors may prove more likely to modulate ErbB receptor behaviour, and knowledge of the precise HER2–Herceptin contacts may allow design or selection of Herceptin variants with higher affinity and improved therapeutic properties. The Herceptin-binding site encompasses a pocket-like structure homologous to the domain II loop-binding pocket of other ErbB receptors, which may accommodate binding of small, perhaps peptide-based molecules with reasonable affinity. Multivalent forms of such a compound may provide a new weapon against HER2-overexpressing cancers. Furthermore, if the exposed domain II loop of HER2 proves important for mediating interactions of HER2 with itself or with other ErbB receptors, agents targeting this loop, for example small peptides or peptide analogues derived from this region, may inhibit such interactions. □

## Methods

### Expression of rat sHER2

Residues 22–635 of rat HER2 followed by a 7-His tag were expressed in BW5147.G.1.4 cells using HI-GET technology<sup>21</sup>. sHER2 was purified using Ni-NTA resin, chromatographed on a gel-filtration column, deglycosylated with PNGase F, re-chromatographed on a gel-filtration column, concentrated to 5 mg ml<sup>−1</sup>, and dialysed against 10 mM Tris, pH 7.5. Crystals of sHER2 were grown by the hanging-drop vapour diffusion method using drops with equal parts protein solution and 15–30% PEG 4000, 50 mM sodium citrate, pH 5.4, and 10 mM EDTA. Crystals grew in space group *P*<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> with cell dimensions *a* = 130.87, *b* = 116.40, *c* = 55.37 Å.

### Expression of human sHER2

Residues 1–631 of human HER2 were expressed as a fusion protein with human growth hormone in Lec1 cells using the pSGHV0 expression vector<sup>22</sup>. Human sHER2 fusion protein was purified using Ni-NTA resin, cleaved by TEV protease, and re-chromatographed on a gel filtration column. Purified sHER2 was dialysed into 10 mM sodium citrate, pH 5.5 and 50 mM NaCl, treated with endoglycosidase H,

re-chromatographed on a gel filtration column, concentrated to 3.8 mg ml<sup>−1</sup>, and dialysed against 10 mM Tris, pH 7.5, and 50 mM NaCl.

### The sHER2–Herceptin Fab complex

Herceptin (also known as trastuzumab) was purchased from Genentech. Ten milligrams of Herceptin were cleaved with papain, and the Fab fragment was purified by anion exchange chromatography using a Resource-Q column. Human sHER2 was mixed with an excess of the Herceptin Fab, and the complex was purified by gel-filtration chromatography. This complex was dialysed against 10 mM Tris, pH 7.5, and concentrated to 4 mg ml<sup>−1</sup>. Crystals were grown by the hanging-drop vapour diffusion method using equal parts protein solution and 20–30% PEG 5000 MME, 0.1 M MES, pH 6.5, and 0.2 M ammonium sulphate. Crystals grew in space group *P*<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> with cell dimensions *a* = 66.17, *b* = 109.77, *c* = 175.14 Å.

### Structure determination and refinement

Diffraction data were collected at beamlines X4A and X25 at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. The structures of rat sHER2 and human sHER2 complexed with the Herceptin Fab fragment were determined by molecular replacement using AMORE<sup>23</sup>, CNS<sup>24</sup> and MOLREP<sup>25</sup> with individual domains of sHER3<sup>12</sup> and the A5B7 Fab<sup>13</sup> as search models. Alternate sessions of model building using O<sup>26</sup> and refinement using CNS resulted in placement of nearly all of the sHER2 and Herceptin amino acids. Final rounds of refinement were carried out using TLS refinement as implemented in REFMAC5<sup>27</sup> with anisotropic motion tensors refined for each of the four sHER2 domains.

The final rat sHER2 structure consists of residues 1–101, 113–253 and 259–608, 85 water molecules, and 3 *N*-acetylglucosamine moieties. The final human HER2–Herceptin Fab structure consists of residues 1–101, 111–302, 306–360, 365–580 and 591–607 of sHER2, antibody residues 1–211 and 1–223 for the light and heavy chains, respectively, 79 water molecules, one sulphate ion, and two *N*-acetylglucosamine moieties. All stereochemical parameters analysed by the program PROCHECK<sup>28</sup> are within or exceed normal ranges, and the percentage of residues in the most favoured regions of the Ramachandran plot are 86.8% and 84.8% for rat sHER2 and human sHER2 plus the Herceptin Fab, respectively.

Received 17 September; accepted 17 December 2002; doi:10.1038/nature01392.

- Yarden, Y. & Slivkowsky, M. X. Untangling the ErbB signalling network. *Nature Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
- Olayioye, M. A., Neve, R. M., Lane, H. A. & Hynes, N. E. The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J.* **19**, 3159–3167 (2000).
- Tang, C. K. & Lippman, M. E. in *Hormones and Signaling* (ed. O'Malley, B. W.) 113–165 (Academic, San Diego, 1998).
- Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177–182 (1987).
- Slamon, D. J. *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344**, 783–792 (2001).
- Carpenter, G. Receptors for epidermal growth factor and other polypeptide mitogens. *Annu. Rev. Biochem.* **56**, 881–914 (1987).
- Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **103**, 211–225 (2000).
- Jones, J. T., Akita, R. W. & Slivkowsky, M. X. Binding specificities and affinities of egf domains for ErbB receptors. *FEBS Lett.* **447**, 227–231 (1999).
- Di Fiore, P. P. *et al.* erbB-2 is a potent oncogene when overexpressed in NIH/3T3 cells. *Science* **237**, 178–182 (1987).
- Ogiso, H. *et al.* Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775–787 (2002).
- Garrett, T. P. J. *et al.* Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor  $\alpha$ . *Cell* **110**, 763–773 (2002).
- Ferguson, K. M. *et al.* EGF activates its receptor by relieving auto-inhibition of ectodomain dimerization. *Mol. Cell* (in the press).
- Cho, H. S. & Leahy, D. J. Structure of the extracellular region of HER3 reveals an interdomain tether. *Science* **297**, 1330–1333 (2002).
- Banfield, M. J., King, D. J., Mountain, A. & Brady, R. L. VL:VH domain rotations in engineered antibodies: crystal structures of the Fab fragments from two murine antitumor antibodies and their engineered human constructs. *Proteins* **29**, 161–171 (1997).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Berezov, A. *et al.* Disabling receptor ensembles with rationally designed interface peptidomimetics. *J. Biol. Chem.* **277**, 28330–28339 (2002).
- Harari, D. & Yarden, Y. Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. *Oncogene* **19**, 6102–6114 (2000).
- Slivkowsky, M. X. *et al.* Nonclinical studies addressing the mechanism of action of Trastuzumab (Herceptin). *Semin. Oncol.* **26**, 60–70 (1999).
- Molina, M. A. *et al.* Trastuzumab (Herceptin), a humanized anti-Her2 receptor monoclonal antibody, inhibits basal and activated Her2 ectodomain cleavage in breast cancer cells. *Cancer Res.* **61**, 4744–4749 (2001).
- Burke, C. L., Lemmon, M. A., Coren, B. A., Engelman, D. M. & Stern, D. F. Dimerization of the p185neu transmembrane domain is necessary but not sufficient for transformation. *Oncogene* **14**, 687–696 (1997).
- Denney, D. W. Jr Gene amplification methods. US patent 5,776,746 (1998).
- Leahy, D. J., Dann, C. E., Longo, P., Perman, B. & Ramyar, K. X. A mammalian expression vector for expression and purification of secreted proteins for structural studies. *Protein Expr. Purif.* **20**, 500–506 (2000).
- Navaza, J. AMORE: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).



24. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
25. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
26. Jones, T., Zou, J.-Y., Cowan, S. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
28. Laskowski, R. A. *et al.* PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
29. Carson, M. Ribbons. *Methods Enzymol.* **277**, 493–505 (1997).
30. Kraulis, P. J. A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

**Supplementary Information** accompanies the paper on *Nature's* website  
(<http://www.nature.com/nature>).

**Acknowledgements** We thank C. Ogata and M. Becker for assistance at beamlines X4A and X25, respectively, of NSLS at Brookhaven National Laboratory; A. Ullrich for supplying a human HER2 complementary DNA; P. Longo for technical assistance; T. Garrett, S. Yokoyama and colleagues for supplying preprints in advance of publication; S. Yokoyama for coordinates of the EGF–EGFR complex; M. Lemmon, K. Ferguson, M. Amzel, J. Berg, S. Bouyain and W. Yang for discussion and comments on the manuscript; A. Guarne for help with figures; and N. Davidson for assistance with Herceptin. This work was supported by the NIH and the HHMI.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to D.J.L. (e-mail: dleahy@jhmi.edu). Coordinates and structure factors are available from the Protein Data Bank under accession numbers 1N8Y (rat sHER2) and 1N8Z (human sHER2–Herceptin Fab complex).

## Crystal structure of the specificity domain of ribonuclease P

Andrey S. Krasilnikov\*, Xiaojing Yang\*, Tao Pan† & Alfonso Mondragón\*

\* Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208, USA

† Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA

RNAse P is the only endonuclease responsible for processing the 5' end of transfer RNA by cleaving a precursor and leading to tRNA maturation<sup>1,2</sup>. It contains an RNA component and a protein component and has been identified in all organisms. It was one of the first catalytic RNAs identified<sup>3</sup> and the first that acts as a multiple-turnover enzyme *in vivo*. RNAse P and the ribosome are so far the only two ribozymes known to be conserved in all kingdoms of life. The RNA component of bacterial RNAse P can catalyse pre-tRNA cleavage in the absence of the RNAse P protein *in vitro* and consists of two domains: a specificity domain and a catalytic domain<sup>4,5</sup>. Here we report a 3.15-Å resolution crystal structure of the 154-nucleotide specificity domain of *Bacillus subtilis* RNAse P. The structure reveals the architecture of this domain, the interactions that maintain the overall fold of the molecule, a large non-helical but well-structured module that is conserved in all RNAse P RNA, and the regions that are involved in interactions with the substrate.

Bacterial RNAse P can be subdivided into two major types (A and B) on the basis of their sequence characteristics. The best-characterized RNAse P molecules come from two bacteria, *Escherichia coli* and *B. subtilis*, which are paradigms for the A- and B-type molecules, respectively. The RNA component of bacterial RNAse P (P RNA) consists of 350–450 nucleotides, whereas the protein component (P protein) is a small, basic protein of about 120 amino acids. In *B. subtilis* P RNA, the specificity domain (S domain)

comprises nucleotides 86–239, and the catalytic domain (C domain) comprises the rest of the molecule (Fig. 1a). The S domain alone can bind pre-tRNA directly with micromolar affinity<sup>6</sup>.

The overall structure of the specificity domain is shown in Fig. 1b together with a diagram illustrating the secondary structure of the molecule (Fig. 1c). The S domain consists of several distinct secondary structure modules, which were predicted from the P RNA sequence and by phylogenetic comparison<sup>7,8</sup>. Overall, the structure agrees very well with secondary structure predictions<sup>8</sup>, cross-linking data<sup>9</sup> and Fe(II)-EDTA cleavage protection data<sup>10</sup> (see Supplementary Information). The most salient features of the structure are (1) a junction formed by the stacked P7, P10 and P11 and the stacked P8 and P9 helices; (2) the packing of the P10.1 and P12 helices through a GAAA tetraloop–tetraloop receptor interaction; and (3) an unusually folded module linking P11 and P12 (J11/12–J12/11) (orange in Fig. 1), which contains a large number of universally conserved nucleotides, and is stabilized without canonical Watson–Crick base pairing. Although the S domain forms a well-packed and compact structure, it is important to note that the P11 and P9 helices together with the J11/12–J12/11 module form a clamp-like opening that contains nucleotides involved in pre-tRNA binding and is large enough to accommodate the TΨC stem-loop of a pre-tRNA molecule.

There are two molecules in the crystallographic asymmetric unit (1 and 2) related by a rotation of about 90°. It is unlikely that the crystallographic dimer observed is related to the dimer formed by the P RNA in the presence of the P protein in solution<sup>11</sup>, as the intermolecular interactions observed in the crystal structure would lead to higher-order aggregates. The two molecules in the asymmetric unit are in slightly different conformations (see Supplementary Information). Molecule 1 has a disordered P12 helix and a partly disordered P10.1 helix (between bases A142 and G166). The tetraloop–tetraloop receptor interaction observed in molecule 2 is precluded in molecule 1 by crystal packing. Comparison of the two conformations suggests that the structure is built of relatively rigid structural elements and stabilized by a variety of interactions, such as stacking of helices, stacking of bulged bases, and the tetraloop–tetraloop receptor interaction. The central element of the S domain is a rigid core (red in Fig. 1), which has a practically identical conformation in both molecules in the asymmetric unit (root-mean-square deviation (r.m.s.d.) = 0.63 Å). The core is formed by the continuous stacking of stems P10 and P11, together with the basal portion of the P10.1 stem. The base of the P10.1 stem includes a loop spanning U175 to A179 (see Fig. 2a). The phylogenetically conserved U175 and A179 form a reversed Watson–Crick base pair that stabilizes the loop and serves as the starting base pair for the P10.1 helix, helping to align the P10.1 stem with respect to the core. Two conserved adenosines in the loop, A177 and A178, enter the minor groove of the P7/P10 stack in a region including conserved base pairs G90–C235 and G132–C234. This type of interaction, between conserved adenosines in a loop and conserved G–C base pairs in a helix, has recently been termed an A-minor motif<sup>12</sup>. It includes contacts between the 2'-OH groups of the loop nucleotides with the 2'-OH groups of the helical residues similar to those in a ribose zipper<sup>13</sup>. The first four canonical Watson–Crick pairs in the P10.1 stem complete this portion of the rigid core. The second helix in the core is formed by the stacking of helices P10 and P11. This helix also stacks on stem P7 to form a large, continuous helix (Fig. 1). We do not include P7 in the rigid core because the P7/P10 stack is rather flexible (the angle between P7 and P10 differs by about 10° between molecules 1 and 2, as opposed to the P10/P11 stack in which the two molecules show identical conformations). The P11 stem includes bulged conserved adenosines A229 and A230 that do not form part of the helix.

The P7/P10/P11 helix is part of a large junction that also includes stacked helices P8 and P9. Figure 2b shows the topology of this complex junction. The general stacking of the helices is in agree-

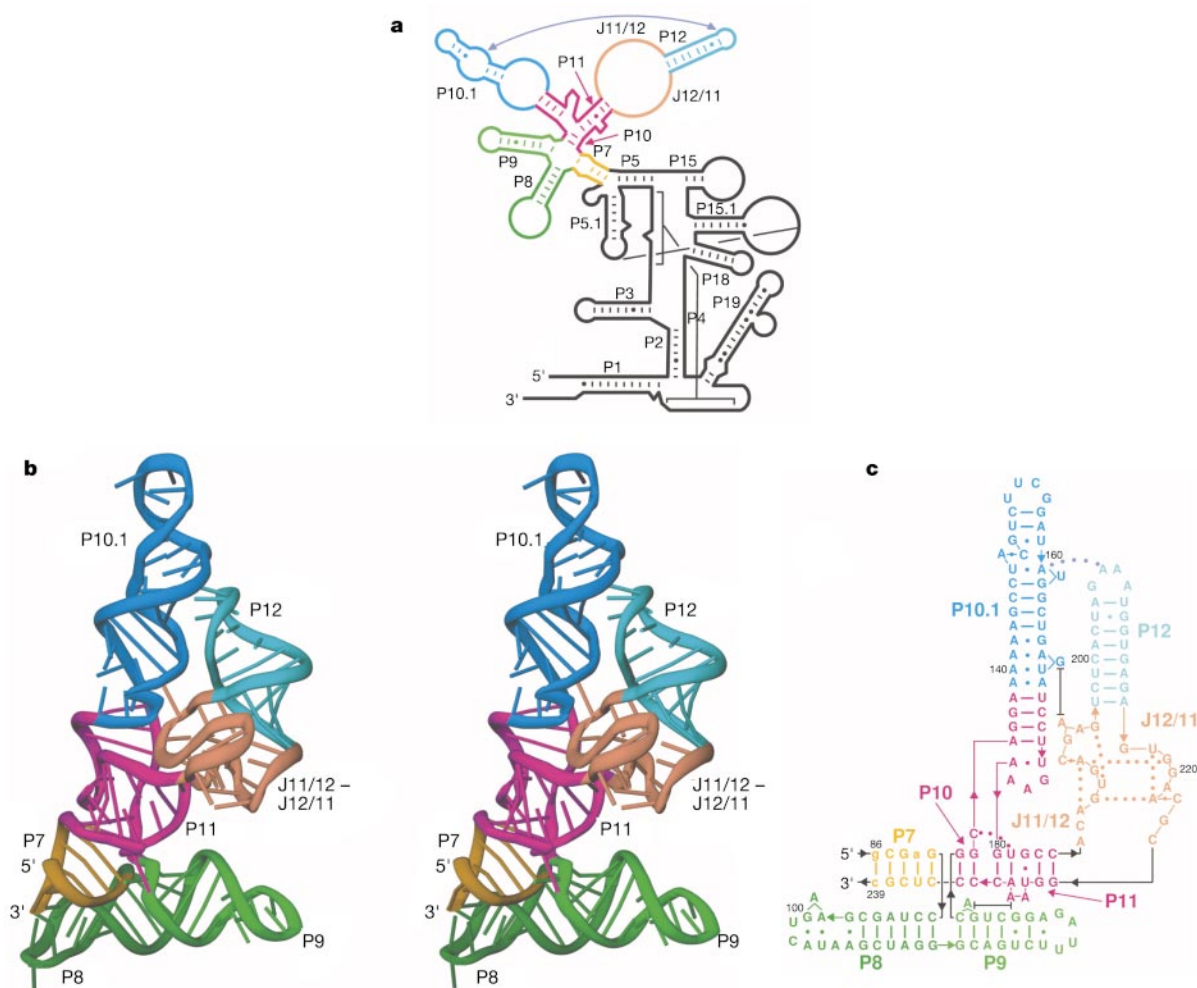


ment with predictions based on modelling studies<sup>8</sup>. Non-Watson–Crick interactions in the region revealed by the crystal structure include the C134•(U181–A231) triplet, the stacking of bases A130 and A230 and the intercalation of A229 between G133 and C134. The last two interactions involve a sharp kink in the phosphate backbone to allow A229 to intercalate between G133 and C134. The backbone then folds back almost 180° so that A230 faces in the opposite direction to A229, and finally another kink resumes the regular helical structure (Fig. 2b, inset). Interestingly, bases involved in these interactions are conserved in bacterial P RNAs, underlining their importance in the correct folding and packing of the molecule as well as their involvement in substrate binding. The stacked bases A130 and A230 are especially interesting because they protrude from the body of the molecule and are protected from chemical modifications by the presence of substrate, both in *B. subtilis*<sup>14</sup> and in *E. coli*<sup>15</sup>. The stacked helix P8/P9 is capped on the P8 side by a large loop stabilized by a sheared base pair G97•A106, a reversed Hoogsteen base pair A98•U104 and a *trans* Sugar-edge/Watson–Crick base pair G100•A103.

Helix P11 is connected to a module formed by two large internal loops (J11/12–J12/11) consisting of A185–G196 and G217–C225 (orange in Fig. 1). These loops have a very similar structure in both molecules in the asymmetric unit (r.m.s.d. = 1.27 Å), although their relative position with respect to P11 is somewhat different,

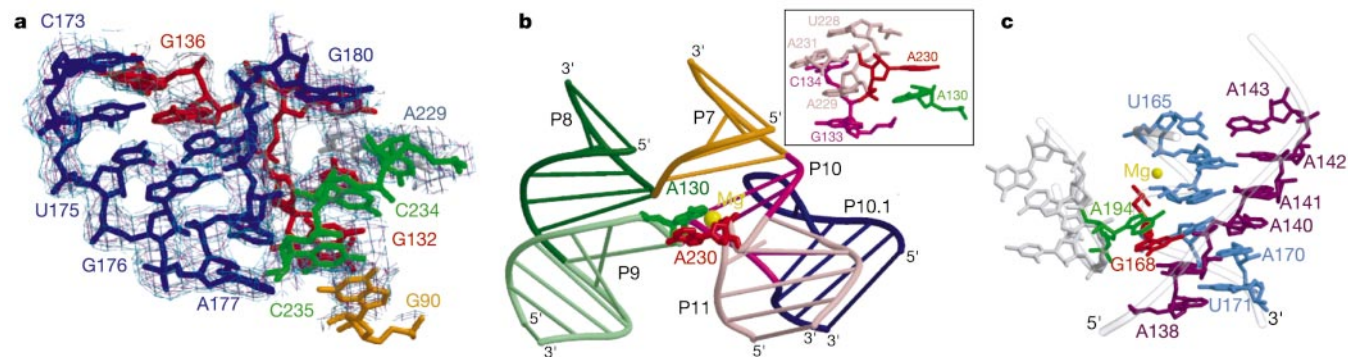
by about 10° (see Supplementary Information). The two chains in the module form a very closely packed structure (Fig. 3a) stabilized by an intricate network of interactions (major interactions are shown in Fig. 1c). There are no canonical Watson–Crick base pairs in this region, creating a large and stable structure without helical base pairing. The J12/11 strand (red in Fig. 3a) forms a loop that starts at the 3′ end of P12 and is stabilized by a reversed Hoogsteen pair, U218•A222. It then folds back to form a side-by-side base pair between adjacent nucleotides A222 and C223. The arrangement of these adjacent A and C nucleotides could be described as a dinucleotide platform, similar to an AA platform<sup>16</sup>. The J11/12 strand (blue in Fig. 3a) forms an S-shaped curve, stabilized by interactions between U189•G196 and A187•A191. Adjacent nucleotides A191 and C192 form another AC platform. In both cases, AC platforms are associated with a sharp, almost 180°, turn in the phosphate backbone.

The central part of P10.1, just above the central core region, contains four non-Watson–Crick base pairs with a bulged G168, as was predicted<sup>8</sup>, and folds into a cross-strand A-stack joined to a bulged G motif<sup>17</sup> (Fig. 2c). The bulged G168 stacks on A194, which sticks out of the J11/12–J12/11 region. Stem P12 is a practically perfect eight-base-pair A-type helix, capped by a GAAA tetraloop. The tetraloop interacts with the tetraloop receptor in P10.1, as has been suggested previously<sup>18</sup>. The sequence of the tetraloop receptor,



**Figure 1** Structure of the S domain of *B. subtilis* RNase P RNA. **a**, Secondary structure of the entire P RNA. The P RNA consists of two well-defined domains, a specificity or S domain (shown in colours) and a catalytic or C domain (black). The diagram shows the secondary structure prediction and nomenclature according to ref. 30. Yellow, the P7 stem; green, the P8 and P9 stems; red, the central rigid core; blue, the distal part of

P10.1; cyan, the P12 stem; orange, the J11/12–J12/11 module. **b**, Stereo ribbon diagram of the structure of the S domain. **c**, Sequence and secondary-structure diagram of the S domain. The nucleotides shown in lower-case letters correspond to mutations introduced in the crystallized fragment. Lines terminated by bars correspond to stacking interactions.



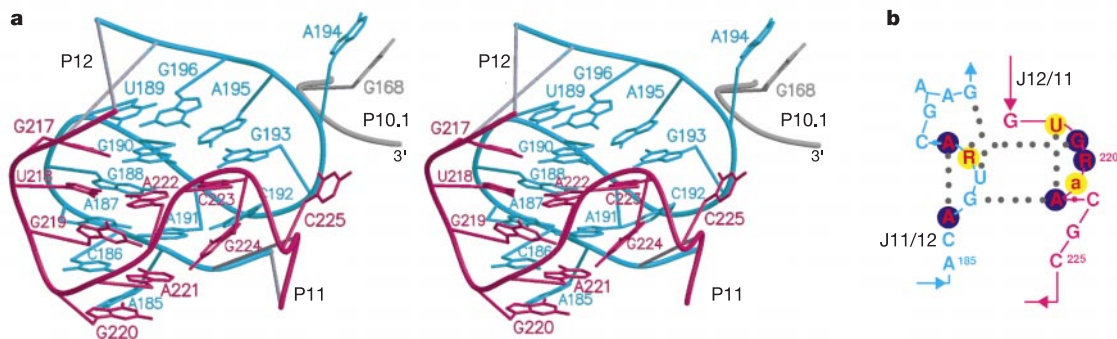
**Figure 2** The central junction and the P10.1 helix. **a**, The loop at the basal part of the P10.1 stem interacts with the P7/P10 stack. Bases A177 and A178 in the loop interact with G–C base pairs at the junction of the P7 and P10 stems, forming an A-minor motif<sup>12</sup>. The solvent flattened experimental electron density map is shown at the 1.0 $\sigma$  and 1.5 $\sigma$  levels. **b**, The central part of the S-domain structure contains a junction bringing

stems P7–P11 together. Two adenosines in the P11 stem bulge out; one of them, A230, stacks on A130 of the P9 stem. The other adenosine, A229, intercalates between G133 and C134 (inset). **c**, The central section of the P10.1 stem forms a cross-strand A-stack attached to a bulged G motif. Bulged G168 (red) stacks on A194 (green), which bulges out from J11/12 segment.

positioned in the upper part of P10.1, is the same as that found in group I introns<sup>19</sup>, with one exception: the AA platform in the group I intron structure<sup>13,16</sup> is replaced with an AC platform in the *B. subtilis* S-domain tetraloop receptor, which is consistent with selection experiments *in vitro* indicating that AA and AC platforms in tetraloop receptors are interchangeable<sup>20</sup>. This replacement does not cause a substantial change in the geometry of the tetraloop–tetraloop receptor complex (r.m.s.d. = 0.91 Å).

The structure of the S domain of *B. subtilis* RNase P provides the basis for understanding the structure of other bacterial RNase P molecules, because sequence analysis suggests that the S domains of all bacterial P RNAs have a common core that comprises stems P7–P11 plus the J11/12–J12/11 module<sup>21</sup>. Alignment and comparison of 30 sequences of the S domain from B-type RNase P molecules<sup>22</sup>, guided by the crystal structure of the *B. subtilis* S domain, clearly show that many of the conserved nucleotides are involved in interactions that are crucial for RNA folding (see Supplementary Information). The crystal structure of the S domain also allows important observations to be made for the S domains from A-type bacterial RNase P RNA, for which more than 300 sequences are known<sup>22</sup>. The major difference between the S domains of the A- and B-type bacterial RNase P molecules is the absence of the P10.1 stem and the addition of the P13 and P14 stems in the

A-type RNase P molecules. It has been suggested previously<sup>4</sup> that the P13/P14 stems in A-type RNase P RNA, modelled as stacked helices<sup>8</sup>, could have a similar role to the P10.1 stem in B-type RNase P RNA. Structure-based alignment of the A-type RNase P RNA sequences suggests that this could indeed be so. The position where the P13/P14 stems are inserted is located in the J12/11 strand. At this position in the S-domain structure the chain forms an exposed surface loop that faces the P10.1 helix. From this position in the J12/11 module, the P10.1 stem is only about 13 Å away. Therefore, P13/P14 could in principle form a long helix positioned in a similar orientation as the P10.1 stem. If the ends of the P13/P14 stems interact with the P12 and P8 stems, as suggested previously<sup>7,8</sup>, the P13/P14 stem would indeed have an equivalent structural role to the P10.1 stem, giving the A- and B-type S domains very similar overall structures. The junction formed by stems P7–P11 is present in both bacterial types. In the B-type RNase P RNA, the P10.1 stem is inserted in the P10/P11 interface. Although A-type RNase P RNA does not have the P10.1 stem, there is a small bulged loop at the same position that might have an equivalent role in altering the P10/P11 interface. The J11/12–J12/11 module contains two of the five completely conserved regions in all RNase P molecules (Fig. 3b), and the structure of this module is expected to be very similar in all P RNAs from Bacteria, Archaea and Eukarya<sup>21,23</sup>.



**Figure 3** The J11/12–J12/11 module. **a**, The J12/11 segment (217–225) forms an internal loop (218–222) with G220 at its tip. The segment continues as the backbone makes a turn, forming an AC platform (A222–C223) and then joins the P11 stem. The J11/12 segment (185–196) is S-shaped and starts with a loop (nucleotides 187–191) which is structurally similar to the 218–222 loop (r.m.s.d. = 1.38 Å). Then the segment changes direction and forms an AC platform (A191–C192). The remaining part of the

segment forms a smooth transition to the P12 stem. The A187–A191, U218–A222 and U175–A179 loops form a structural motif that is abundant in large RNA molecules (A.S.K. and A.M., unpublished observations). **b**, Conserved nucleotides in the J11/12–J12/11 module (circled). Sequence data<sup>22</sup> from Bacteria, Archaea and Eukarya were compared. Nucleotides in the yellow circle (lower-case letters), in the yellow circle (capitals) and circled in blue are at least 90%, 95% and 99% conserved, respectively.

One of the most remarkable interactions observed in the structure involves the stacking of bases that bulge out of the P9 and P11 stems. A pair of adenosines bulge out of the P11 helix, and one of them (A230 in *B. subtilis*) faces the P9 stem and stacks with an adenosine (A130) bulging out of the P9 stem (Fig. 2b). These nucleotides are involved in direct interactions with the pre-tRNA substrate<sup>14,15</sup>. The *B. subtilis* P RNA contacts at least three 2'-OH groups in the T $\Psi$ C stem and T-loop of pre-tRNA; for example, A230 directly interacts with the 2'-OH of nucleotide 62 in the T $\Psi$ C stem<sup>24,25</sup>. Consistent with this interaction is the observation that A230 in the structure of the S domain faces the solvent and the potential hydrogen acceptor, the N1 nitrogen of the base, is completely exposed. The bases of A130 and G220 are also exposed to the solvent and poised to interact with the substrate. Several of the contact sites in the pre-tRNA are in the same groove as the cleavage site but on the opposite face of the helix. Binding to one face of the T $\Psi$ C/acceptor stem by the S domain would therefore completely expose the cleavage site on the other side to be approached by the C domain.

The positions of the three nucleotides (A130, A230 and G220) directly involved in interactions with the pre-tRNA substrate are highlighted in Fig. 4. They belong to the region that has a clamp-like shape and forms a large opening that is more than 15 Å at its widest point. The bases of A230 and G220 are about 19 Å away, which is commensurate with the diameter of an RNA helix. The size of the aperture created by the S domain and the positions of the nucleotides involved in direct contacts make it tantalizing to suggest that the T $\Psi$ C stem and possibly the T-loop of pre-tRNA enter this region. Photoaffinity cross-linking experiments have shown that in *E. coli* RNase P, tRNA interacts with nucleotides in the P8–P9 stems<sup>26</sup>, in agreement with the notion that the pre-tRNA is bound in this region. There is not enough experimental information to position the pre-tRNA in the S-domain structure unambiguously, but the size and shape of this region are ideal for interactions with the substrate. The high degree of sequence conservation in this region also suggests that all bacterial RNase P RNAs interact similarly with the substrate.

The regions protected from Fe(II)-EDTA cleavage in the presence of the P protein<sup>5</sup> and under conditions leading to holoenzyme

dimerization *in vitro*<sup>11</sup> are also highlighted in Fig. 4. It is likely that only some of these regions are involved in binding to the P protein or in dimer contacts, and that the protection data also reflect changes in conformation due to the binding of P protein, contacts between P RNA and P RNA in the holoenzyme, or both. Nevertheless, it is striking that the P protein probably affects the structure in close proximity to the pre-tRNA-binding region, suggesting a direct coupling between pre-tRNA binding and P protein action.

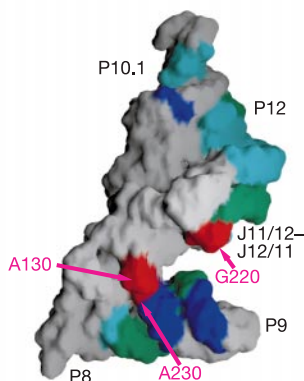
The structure of the S domain of *B. subtilis* RNase P provides a molecular framework for studying the interactions that must occur between tRNA and the ribozyme during pre-tRNA processing. The structure is consistent with the available biochemical data for bacterial RNase P RNA and extends our understanding of RNase P structure across all taxonomic kingdoms. Furthermore, as an important addition to the still limited number of large RNAs of known structure, it advances our knowledge of the general principles of RNA structure and packing. □

## Methods

The crystal structure of the S domain of *B. subtilis* RNase P was solved to 3.15 Å resolution from a two-wavelength multiwavelength anomalous diffraction experiment<sup>27</sup> with a Pb derivative. Refinement was done with Refmac5 (ref. 28) and CNS<sup>29</sup> without the use of non-crystallographic symmetry. The final  $R_{\text{factor}}$  and  $R_{\text{free}}$  values were 28% and 30.7%, respectively. The r.m.s.d. for bond lengths, angles and average  $B$  factor are 0.008 Å, 1.29° and 94.2 Å<sup>2</sup>. The final model contains all RNA atoms except disordered bases 143–165 and 201–212 in molecule 1 and 121–124 in molecule 2. Only the phosphate backbone could be traced for bases 99 and 101 in both molecules. A combined model representing the entire S domain was built mostly from molecule 2, with bases 86–133 and 234–239 from molecule 1 and was used for all figures. Figures were prepared with Ribbons, Molscript, Bobscript and Grasp. For additional experimental details and references, see Supplementary Information.

Received 7 October; accepted 16 December 2002; doi:10.1038/nature01386.

- Altman, S. & Kirsebom, L. A. in *The RNA World* (eds Gesteland, R. E., Cech, T. R. & Atkins, J. F.) 351–380 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
- Frank, D. N. & Pace, N. R. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.* **67**, 153–180 (1998).
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
- Loria, A. & Pan, T. Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA* **2**, 551–563 (1996).
- Pan, T. Higher order folding and domain analysis of the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry* **34**, 902–909 (1995).
- Qin, H., Sosnick, T. R. & Pan, T. Modular construction of a tertiary RNA structure: the specificity domain of the *Bacillus subtilis* RNase P RNA. *Biochemistry* **40**, 11202–11210 (2001).
- Brown, J. W. *et al.* Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements. *Proc. Natl Acad. Sci. USA* **93**, 3001–3006 (1996).
- Massire, C., Jaeger, L. & Westhof, E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* **279**, 773–793 (1998).
- Chen, J. L., Nolan, J. M., Harris, M. E. & Pace, N. R. Comparative photocross-linking analysis of the tertiary structures of *Escherichia coli* and *Bacillus subtilis* RNase P RNAs. *EMBO J.* **17**, 1515–1525 (1998).
- Pan, T. Novel RNA substrates for the ribozyme from *Bacillus subtilis* ribonuclease P identified by *in vitro* selection. *Biochemistry* **34**, 8458–8464 (1995).
- Barrera, A. *et al.* Dimeric and monomeric *Bacillus subtilis* RNase P holoenzyme in the absence and presence of pre-tRNA substrates. *Biochemistry* **41**, 12986–12994 (2002).
- Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA* **98**, 4899–4903 (2001).
- Cate, J. H. *et al.* Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1685 (1996).
- Odell, L., Huang, V., Jakacka, M. & Pan, T. Interaction of structural modules in substrate binding by the ribozyme from *Bacillus subtilis* RNase P. *Nucleic Acids Res.* **26**, 3717–3723 (1998).
- LaGrande, T. E., Huttenhofer, A., Noller, H. F. & Pace, N. R. Phylogenetic comparative chemical footprint analysis of the interaction between ribonuclease P RNA and tRNA. *EMBO J.* **13**, 3945–3952 (1994).
- Cate, J. H. *et al.* RNA tertiary structure mediated by adenosine platforms. *Science* **273**, 1696–1699 (1996).
- Moore, P. B. Structural motifs in RNA. *Annu. Rev. Biochem.* **68**, 287–300 (1999).
- Tanner, M. A. & Cech, T. R. An important RNA tertiary interaction of group I and group II introns is implicated in gram-positive RNase P RNAs. *RNA* **1**, 349–350 (1995).
- Costa, M. & Michel, F. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276–1285 (1995).
- Costa, M. & Michel, F. Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution. *EMBO J.* **16**, 3289–3302 (1997).
- Chen, J.-L. & Pace, N. R. Identification of the universally conserved core of ribonuclease P RNA. *RNA* **3**, 557–560 (1997).
- Brown, J. W. The Ribonuclease P Database. *Nucleic Acids Res.* **27**, 314 (1999).
- Frank, D. N., Adamidi, C., Ehringer, M. A., Pitulle, C. & Pace, N. R. Phylogenetic-comparative analysis of the eukaryal ribonuclease P RNA. *RNA* **6**, 1895–1904 (2000).



**Figure 4** Surface representation of the S-domain structure. Three nucleotides (A130, A230 and G220) protected from chemical modifications by the substrate in *B. subtilis* RNase P<sup>14</sup> are shown in red. The position and distance between the nucleotides suggest that the clamp could serve to semi-enclose a large region of the substrate, for example the T $\Psi$ C stem. Regions protected from Fe(II)-EDTA cleavage owing to dimerization of the RNase P holoenzyme<sup>11</sup> (light blue), the presence of the protein<sup>5</sup> (dark blue) and the overlap of the two (green) are also shown. The protection experiments might not reflect the exact position of the protein or dimerization regions because conformational changes can affect the results, but the data serve to indicate that the dimerization and protein interaction regions are all located in the same face of the S domain and in proximity to the pre-tRNA-binding site.



24. Pan, T., Loria, A. & Zhong, K. Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. *Proc. Natl Acad. Sci. USA* **92**, 12510–12514 (1995).
25. Loria, A. & Pan, T. Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry* **36**, 6317–6325 (1997).
26. Nolan, J. M., Burke, D. H. & Pace, N. R. Circularly permuted tRNAs as specific photoaffinity probes of ribonuclease P RNA structure. *Science* **261**, 762–765 (1993).
27. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
29. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
30. Haas, E. S., Banta, A. B., Harris, J. K., Pace, N. R. & Brown, J. W. Structure and evolution of ribonuclease P RNA in Gram-positive bacteria. *Nucleic Acids Res.* **24**, 4775–4782 (1996).

**Supplementary Information** accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

**Acknowledgements** We thank X. Liu and Y. Xiao for technical assistance, A. Changela, H. Feinberg, V. Grum and members of DND–CAT for help with data collection, and A. Changela, C. Correll, V. Grum, E. Sontheimer, B. Taneja and J. Wedekind for comments and suggestions. Research was supported by the NIH (to A.M.) and an NIH NRSA Fellowship to A.K. Support from the R.H. Lurie Cancer Center of Northwestern University to the Structural Biology Center is acknowledged. Portions of this work were performed at the DuPont–Northwestern–Dow Collaborative Access Team (DND–CAT) Synchrotron Research Center at the Advanced Photon Source (APS) and at the Stanford Synchrotron Radiation Laboratory (SSRL). DND–CAT is supported by DuPont, Dow and the NSF, and use of the APS is supported by the DOE. SSRL is operated by the DOE, Office of Basic Energy Sciences. The SSRL Biotechnology Program is supported by the NIH and the DOE.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to A.M. (e-mail: a-mondragon@northwestern.edu). Coordinates have been deposited in the Protein Data Bank under the accession code 1NBS.

## retractions

### A cytosolic catalase is needed to extend adult lifespan in *C. elegans* *daf-1* and *clk-1* mutants

J. Taub, J. F. Lau, C. Ma, J. H. Hahn, R. Hoque, J. Rothblatt & M. Chalfie

*Nature* **399**, 162–166 (1999).

We no longer have confidence in our observations associating a reduction in adult lifespan with a putative mutation in the *Caenorhabditis elegans* catalase gene *ctl-1* and therefore retract this paper. With the assistance of J. Liang and C. Keller, we have confirmed that *C. elegans* has multiple catalase genes (actually three in tandem) and that the original strain, TU1061, has decreased transcription of *ctl-1* messenger RNA. However, we have also found several errors, one identifying a single nucleotide deletion as the defect in the putative *ctl-1* mutation and others in the identification of strains carrying mutations in multiple genes. In particular, we have not seen the expected reduction in *ctl-1* mRNA in other

strains tested. The longevity results obtained with these strains are therefore meaningless. We are grateful to our colleagues, particularly C. Kenyon and M. Crowder, for conveying to us their concerns about our results. □

## Metal–insulator transition in chains with correlated disorder

Pedro Carpena, Pedro Bernaola-Galván, Plamen Ch. Ivanov & H. Eugene Stanley

*Nature* **418**, 955–959 (2002).

This Letter reported numerical simulations of one-dimensional disordered binary systems, and found a threshold value for the exponent characterizing the long-range power-law correlations of the system. Below this threshold, the system behaves as an insulator and above it, in the thermodynamic limit, the system behaves as a conductor. Unfortunately, we have now found that this observation was a consequence of the algorithm used to generate long-range correlations in binary chains, because above the threshold value of the exponent only a finite number of segments of atoms of the same type (A or B) exists, even in the thermodynamic limit of an infinitely large system. Thus, the system studied was not truly disordered. As a result, what we observed at the critical threshold value for the correlation exponent was not a transition from insulator to metal behaviour in a disordered system (as reported), but a transition from a disordered to an ordered system. For this reason, the authors retract the claim of a metal–insulator transition in the infinite binary chain with correlated disorder. The results are still valid that relate to the behaviour of a binary chain below the critical threshold value of the correlation exponent, and to large but finite system sizes (as found in the DNA example discussed in the Letter).

We thank L. Hufnagel and T. Geisel for drawing this to our attention. □

## erratum

### Wave-like properties of solar supergranulation

L. Gizon, T. L. Duvall Jr & J. Schou

*Nature* **421**, 43–44 (2003).

In Fig. 1, the units of frequency should be microhertz (μHz), not millihertz (mHz). In the US-printed issues, Fig. 3b appeared blurred. □

## Contacts

**Publisher:** Ben Crowe

**Editor:** Paul Smaglik

**Marketing Manager:** David Bowen

### European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**Senior European Sales Manager:**  
Nevin Bayoumi (4978)

### UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

### Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

**Scandinavia:** Silje Opstrup (4994)

### France/ Belgium:

Amelie Pequignot (4974)

### Production Manager:

 Billie Franklin

To send materials use London  
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### International

#### Advertising Coordinator:

Hind Berrada (4935)

#### Naturejobs web development:

Tom Hancock

#### Naturejobs online production:

Ben Lund

### European Satellite Office

#### Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mail: [p.phelan@nature.com](mailto:p.phelan@nature.com)

[o.wulffen@nature.com](mailto:o.wulffen@nature.com)

### US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

### US Sales Manager:

 Peter Bless

### US Advertising Coordinator:

Linda Adam

### Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakutamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

### Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: [h.watanabe@naturejp.com](mailto:h.watanabe@naturejp.com)

# naturejobs

## A moving story

In the winter of 2001, it was Peter Schlögelhofer's turn to face the 'postdoc dilemma'. At the time, he was a graduate student in plant molecular biology at the University of Vienna and was seeking a postdoc position in the United States. He wanted to find the best scientific fit for his background — but he also needed the post to offer him a good quality of life on the limited income he was set to receive. Schlögelhofer was fortunate enough to receive several offers, which he had to weigh carefully.

The first position was at a prestigious West Coast university. It was tempting scientifically, but the housing costs were far too high. A second offer came from an East Coast institution, which allowed for reasonable housing, but the project wasn't a terribly good fit. The third offer was from a state university, which lacked a 'designer' name but which offered a project that Schlögelhofer was enthusiastic about, as well as housing that wouldn't stretch his stipend.

Schlögelhofer's reasoning shows how house prices can affect a postdoc's decision process — perhaps more than many realize or care to admit. Many institutions in expensive cities are recognizing how much of a burden housing can place on prospective fellows, and they are beginning to address the problem. But as yet there is no uniform approach (see page 766).

Just as Schlögelhofer was about to settle on his third option — he thought the project's quality, the adviser's reputation and the reasonable rent trumped the institutional cachet of the first two — he changed his mind. He decided to trade his scientific prospects abroad for a romantic one at home. Fortunately, it all worked out in the end. He landed a temporary position at the University of Vienna, complete with his own lab and grant-writing responsibilities. And he doesn't have to worry about housing — unless he decides to leave Vienna for a permanent position when his present contract ends in two years' time.

### Paul Smaglik

*Naturejobs* editor



## Contents

### POSTDOCS & STUDENTS

Hot property p766

### MOVERS

A fourth continent for physicist;  
genomics institute gets new  
head; Italy's loss is France's gain;  
and more p768

### WWW.NATUREJOBS.COM

Career centre  
Information on the  
scientific job market

### FOCUS

### SPOTLIGHT

### RECRUITMENT

### SCIENTIFIC ANNOUNCEMENTS

### SCIENTIFIC EVENTS

# Location, location, location

When a postdoc has a choice about where to go for a fellowship, how much do non-scientific factors like cost of living weigh on the final decision?

Karen Kreeger investigates.



Without subsidies, Herman Wijnen (inset) believes it would be hard to find convenient properties in New York.

When asked why they chose a particular fellowship, postdocs usually give scientific reasons: the reputation of the institution, the quality of the lab, the track record of the principal investigator. Those factors are paramount, as they should be. But when pressed about how they are managing financially, most of them will say that their main concern is housing.

Postdocs in the United States encounter a wide range of answers to the housing question from different institutions, even among those with equal prestige. In expensive parts of the east and west coasts, institutions have to compete for the best postdocs, and have come up with different ways of attracting them. Some offer more pay, others provide subsidies, and some are building affordable housing themselves.

## ON THE EAST SIDE

Rockefeller University eases the high cost of living in New York City by helping with housing costs — as well as with childcare for those who need it.

Postdoc Herman Wijnen says that he chose Rockefeller for the quality of its science. But the university's policy of heavily subsidizing postdocs' housing is a bonus, as it helps him to afford an apartment near the lab. "We're on the Upper East Side, the silk-stockings district, and rents would be totally unaffordable here otherwise," says Wijnen.

Postdocs living in the university's own Manhattan apartments pay 10% less than the standard rent, which is already less than the local average. Kevin O'Donovan, another Rockefeller postdoc, pays just over \$1,300 per month for a one-bedroom apartment. "The going rate in the neighbourhood for an apartment my size would be at least \$2,500–3,000," he says. Most of the Rockefeller housing is within a ten-minute walk of the university.

Even with such subsidies, it is not easy to live in New York on a postdoc's salary. Rockefeller postdoc Allan Coop spends at least half of his after-tax salary

on rent. Coop's wife is unable to work because of visa restrictions, so both of them have to live on his salary. "In fact, it's cost me more to be here than I have earned in the four years since I arrived," he says. "The only way to rationalize that is to say that the whole exercise is an investment in one's future."

Postdocs make ends meet by using public transport, sharing accommodation — and running up debts. Despite all the subsidies, there is still a housing shortage, especially as some institutions grow. The Memorial Sloan-Kettering Cancer Center (MSKCC), also in New York City, is undergoing a major expansion over the next few years. To keep pace with its housing needs, it is building a 270-unit apartment building on Roosevelt Island, just off Manhattan, for professional staff, faculty members, students and technicians, due to open this spring. Thomas Kelly, director of the MSKCC's research arm, the Sloan-Kettering Institute, says that the rents, as yet unset, will be below market value.

But New York is not the only address with problems. Elizabeth Courtenay, a postdoctoral fellow in microbiology at the Massachusetts Institute of Technology (MIT) in Cambridge, says that she looked for a place at either the University of California, Berkeley, or MIT, both in areas known for their high cost of living. Courtenay took matters into her own hands and asked her mentor for a higher salary than was offered at the time under guidelines by the National Institutes of Health for starting postdocs. "I negotiated to start at \$31,000 because I didn't think I could make ends meet on \$28,000," she says, although she recognizes that not all labs can afford to do that. Even so, paying \$1,500 a month for an apartment means that 75% of her income goes on living expenses — but living ten minutes' walk from work means that she does not need a car.

Public transport played a major role in the housing choice made by Mary Stewart, a postdoc at MIT's Center for Cancer Research, and her husband. Having moved to Boston to be near her family, they share a car



those with children or for a foreign postdoc whose spouse is unable to work because of visa constraints.

"If you'd like to live close to Scripps, which most people do in their first few years here, a one-bedroom apartment is at least \$1,100 in this part of San Diego," says Frosst. Two-bedroom apartments cost about \$1,400–1,500 a month. Rents tend to decrease with distance from Scripps. Frosst lives about 15 miles south of the major research institutes, closer to the city centre, where a one-bedroom apartment costs \$800–900. "I can squeak by," she says. "If you don't want to live with multiple roommates you have to compromise." For her that means a longer commute.

Unfortunately, San Diego's public-transport system is poor, unlike those of San Francisco, Boston and New York, where postdocs get by quite easily without a car.

#### IN BETWEEN

Major cities on both coasts may have renowned centres of science, but they are expensive for anyone, let alone postdocs on small salaries. Smaller

towns and cities may offer equally good research facilities but without the high costs.

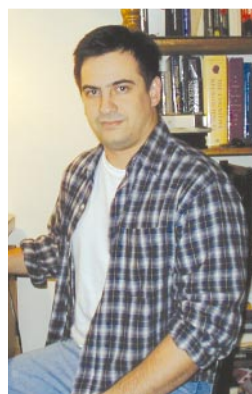
"I think there's no question that there's a significant difference, particularly if you're in a position to buy a house or have small children," says William Snider, director of the Neuroscience Center at the University of North Carolina in Chapel Hill. He says that he finds it easy to recruit postdocs, partly because of the reasonable cost of living, good lifestyle and abundant opportunities.

Nearby Durham has the same advantages. "I think in the \$600 range would be a fair estimate for a one-bedroom apartment," says Susan Murphy, a postdoc at Duke University in Durham and chair of the advocacy committee for the Duke University Postdoctoral Association. In a recent survey conducted by her group, single people reported paying an average of \$645 a month rent, with a range of \$285–1,300; this included a few who said they shared accommodation. Married postdocs paid \$400–2,000 rent, with an average of \$763.

David Pettigrew, a postdoc at the University of Texas Medical Center in Houston, says that he received job offers from Cornell University in Ithaca, New York; Washington University in St Louis; and an institute in Cincinnati. "Even though there was some variance in the salaries that were offered, it did not have a large impact on my decision," he says. He chose to stay in Texas because of the research project offered by his mentor. Pettigrew says that about 40% of his salary goes on rent and he doesn't run a car, taking a bus to work instead.

Although some institutions in expensive areas voluntarily offer subsidized housing or extra money, many postdocs consider that grant-giving institutions should take some of the responsibility. That happens in Britain, where postdocs who work in London receive an "abundant supplement" to offset the cost of living, says Wijnen. He would like to see the same thing happen in the United States. "It wouldn't be unreasonable for stipend levels to be adjusted a little bit to reflect the difference," he says. ■

Karen Kreeger is a freelance science writer based in Philadelphia.



**Difficult decisions:** both Phyllis Frosst and David Pettigrew (inset) agree that housing and the cost of living affect the career choices made by postdocs.

— expensive to insure in the Boston area — and rent a two-bedroom apartment costing \$1,600 a month in Somerville, about 30 minutes' subway ride from the campus. Somerville is a desirable area, and being close to a subway station means that rents are high. "We know we're paying for it, but we're doing it on purpose for convenience," Stewart explains.

#### WAY OUT WEST

Instead of subsidizing accommodation, the Gladstone Institute at the University of California, San Francisco, adds a 10% housing allowance to postdocs' already higher-than-average salaries and provides \$1,500 for moving expenses. This is useful because, as Gladstone postdoc Luke Esposito notes: "San Francisco is expensive." The city exerts two opposite forces on anyone considering a move there, he says: the costs are as daunting as the lifestyle is attractive.

Anyone wanting a place of their own is likely to pay \$850–1,050 a month in rent for a studio or small one-bedroom apartment, he says. Two people could share a larger one-bedroom or a two-bedroom apartment for \$1,500–2,000. But housing costs were not crucial in Esposito's decision to come to San Francisco, he says. "If they didn't have the allowance, I probably still would have come, although it makes us feel good that they're doing something to offset the cost of living," he says. Jeannie Chin, another Gladstone postdoc, agrees: "I really came for this lab. Not only are they doing great science, they're trying to take care of their scientists."

Down the coast, the University of California, San Diego, provides generously subsidized housing. Its faculty members, students and postdocs can rent an apartment for about half the market rate. The neighbouring Scripps Research Institute, on the other hand, doesn't provide any allowances or subsidies.

Scripps postdoc Phyllis Frosst, who is president of its Society of Fellows, notes that in her experience housing does enter into the equation. This is especially true for

## Transitions

X Last December, **Leif Hamø** was appointed chief financial officer at TopoTarget Prolifix, a biotech firm in Copenhagen. Hamø joins the company from Danish credit and finance firm AcceptFinans, where he was chief executive.

X Dendreon of Seattle last month appointed **Israel Rios** as vice-president of clinical affairs. Rios joins the firm from Berlex Laboratories, the US affiliate of Schering, Germany, where he most recently served as vice-president of oncology development.

X **Barbara Domayne-Hayman** is to join London-based Arrow Therapeutics, as commercial director; she will move from Celltech where she is currently senior business-development manager.

X Renovis, a South San Francisco biotech company, last month brought on **John Doyle** as vice-president of finance and chief financial officer; **Jacob Huff** as vice-president of clinical development; **Michael Kelly** as senior director of chemistry; **John Kollins** as vice-president of business development; and **Paula Serbin** as vice-president of human resources.

X Astex Technology, a drug-discovery company based in Cambridge, UK, last month appointed **David Rees** as director of medicinal chemistry. Rees comes from AstraZeneca, in Mölndal, Sweden, where he was director and head of the medicinal chemistry department.

X Last month Oxford drug-discovery company Etiologies appointed **Ian Kent** as chairman. Kent has helped to start several biotech firms including Ardana Biosciences of Edinburgh, UK.

## PHYSICS

**Katepalli Sreenivasan**, professor of physics at the University of Maryland and an expert in fluid dynamics and turbulence, was brought up in India to consider ways of making a contribution to humanity in whatever he was doing. He will have every opportunity to do just that when he becomes director of the Abdus Salam International Centre for Theoretical Physics (ICTP) in Italy next month. Although based in Trieste, the ICTP was founded in 1964 by Abdus Salam with the aim of fostering the growth of science and research in developing countries. "As I am from the developing part of the world it made a lot of sense," Sreenivasan says.

Once he makes the move, Sreenivasan will have been involved in scientific enterprises on four continents. He received his doctorate in aerospace engineering at the Indian Institute of Science in Bangalore, and spent two years as a post-doctoral research fellow in Australia before moving to Johns Hopkins University in Baltimore. He moved from Yale University to Maryland last January.

He has also been adept at moving between disciplines. He was an engineer by training, but took courses in physics and mathematics when he was a student. A few years after he came to the United States, there was a lot of activity in nonlinear dynamics, which led him to study complexity. "Science is about connections, really," Sreenivasan says.

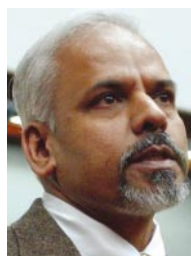
## GENOMICS

**Huntington Willard**, director and president of the research institute of University Hospitals of Cleveland, became director of the Duke University Institute for Genome Sciences and Policy last month. Willard has specialized in understanding how chromosomes function and organize their genes. His group gained international attention in 1997 when it constructed the first human artificial chromosome.

The \$200-million genomics institute was launched in 2000 and is organized into five research areas — the Center for Human Genetics, the Center for Human Disease Models, the Center for Genome Technology, the Center for Bioinformatics and Computational Biology, and the Center for Genome Ethics, Law and Policy. As the institute expands, Willard anticipates hiring 25–40 scientists.

## SPACE

Last month **Giovanni Bignami** left the science directorship of the Italian space agency ASI to take up the directorship of the Laboratory of Space Astrophysics (CESR) in Toulouse, France.



Katepalli Sreenivasan



Huntington Willard



Giovanni Bignami



Robert Brown

Bignami denies that his move is a protest at the Italian government's lack of commitment to basic research, which is how it was interpreted in the Italian press. Although he calls the government's funding policy a "grave mistake", he says that his move is more an endorsement of European science than a condemnation of Italian policy. "France has a tradition of opening key positions to scientists from other countries, and as I take up my new job, I underline the European level at which an institute as significant as the CESR should operate."

Although he is now based in the south of France, Bignami will maintain his chair of astronomy at the University of Pavia in Italy.

## ASTRONOMY

After 33 years with the National Radio Astronomy Observatory (NRAO) in Charlottesville, Virginia, deputy director **Robert Brown** will move in May to become director of the National Astronomy and Ionosphere Center (NAIC). NAIC's main facility is the Arecibo Observatory in Puerto Rico — at 305 metres in diameter it is the world's largest, and most sensitive, single-dish radio telescope. Brown plans to spend about a quarter of his time at the telescope and the rest at NAIC's site at Cornell University in Ithaca, New York.

In recent years, Brown has played a leading role in the international group that is designing and constructing the Atacama Large Millimeter Array (ALMA) observatory in Chile. "As a result of that, I'm aware of the problems that can happen when building big projects in remote locations," says Brown, who adds that even though Arecibo is a radio telescope and ALMA uses an array of collectors, the challenges of managing both facilities from a distance have some similarities. "Astronomers are used to getting data remotely," Brown says.

He aims to increase Arecibo's utility to off-site users. Brown succeeds **Paul Goldsmith**, the J. A. Weeks professor in the physical sciences at Cornell, who stepped down last month to return to full-time research and teaching.

Correction In the supplement *Science & Technology Networks in Scandinavia* (*Nature* **420**, suppl. A21; 2002), Astra Zeneca and The Knut and Alice Wallenberg Foundation are described as having invested in Uppsala-based instrumentation company Gyros. Neither has a financial stake in the firm. *Nature* regrets the error.

## CONTACTING US AT MOVERS

Please send any information on appointments or job moves to:  
 ◆ [naturejobseditor@naturedc.com](mailto:naturejobseditor@naturedc.com).